A general account of argumentation with preferences

Modgil, Sanjay; Prakken, Henry

Published in:
Artificial Intelligence

DOI:
10.1016/j.artint.2012.10.008

Publication date:
2013

Artificial Intelligence 195 (2013) 361–397

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

A general account of argumentation with preferences

Sanjay Modgil [a,*], Henry Prakken [b,c]

[a] Department of Informatics, King's College London, United Kingdom
[b] Department of Information and Computing Sciences, Utrecht University, Netherlands
[c] Faculty of Law, University of Groningen, Netherlands

ARTICLE INFO

Article history:
Received 9 January 2012
Received in revised form 28 October 2012
Accepted 30 October 2012
Available online 5 November 2012

Keywords:
Argumentation
Preferences
Non-monotonic reasoning
Dung

ABSTRACT

This paper builds on the recent *ASPIC*$^+$ formalism, to develop a general framework for argumentation with preferences. We motivate a revised definition of conflict free sets of arguments, adapt *ASPIC*$^+$ to accommodate a broader range of instantiating logics, and show that under some assumptions, the resulting framework satisfies key properties and rationality postulates. We then show that the generalised framework accommodates Tarskian logic instantiations extended with preferences, and then study instantiations of the framework by classical logic approaches to argumentation. We conclude by arguing that *ASPIC*$^+$'s modelling of defeasible inference rules further testifies to the generality of the framework, and then examine and counter recent critiques of Dung's framework and its extensions to accommodate preferences.

1. Introduction

Argumentation is a key topic in the logical study of non-monotonic reasoning and the dialogical study of inter-agent communication [11,45]. Argumentation is a form of reasoning that makes explicit the reasons for the conclusions that are drawn and how conflicts between reasons are resolved. This provides a natural mechanism to handle inconsistent and uncertain information and to resolve conflicts of opinion between intelligent agents. In logical models of non-monotonic reasoning, the argumentation metaphor has proved to overcome some drawbacks of other formalisms. Many of these have a mathematical nature that is remote from how people actually reason, which makes it difficult to understand and trust the behaviour of an intelligent system. The argumentation approach bridges this gap by providing logical formalisms that are rigid enough to be formally studied and implemented, while at the same time being close enough to informal reasoning to be understood by designers and users.

Many theoretical and practical developments build on Dung's seminal theory of abstract argumentation [23]. A Dung *argumentation framework* (*AF*) consists of a conflict-based binary *attack* relation $\mathcal{C}$ over a set of arguments $\mathcal{A}$. The justified arguments are then evaluated based on subsets of $\mathcal{A}$ (*extensions*) defined under a range of semantics. The arguments in an extension are required to not attack each other (extensions are *conflict free*), and attack any argument that in turn attacks an argument in the extension (extensions *reinstate/defend* their contained arguments). Dung's theory has been developed in many directions, including argument game proof theories [34] to determine extension membership of a given argument. Also, several works augment *AF*s with preferences and/or values [5,10,33,40], so that the conflict-free extensions, and so justified arguments, are evaluated only with respect to the successful attacks (*defeats*), where an argument X is said to defeat an argument Y iff X attacks Y and Y is not preferred to X.

The widespread impact of Dung's work can partly be attributed to its level of abstraction. *AF*s can be instantiated by a wide range of logical formalisms; one is free to choose a logical language $\mathcal{L}$ and define what constitutes an argument and

* Corresponding author.
E-mail addresses: sanjay.modgil@kcl.ac.uk (S. Modgil), H.Prakken@uu.nl (H. Prakken).

0004-3702/$ – see front matter
http://dx.doi.org/10.1016/j.artint.2012.10.008

attack between arguments defined by a theory in $\mathcal{L}$. The theory's inferences can then be defined in terms of the conclusions of the theory's justified arguments. Indeed, the inference relations of existing logics, including logic programming and various non-monotonic logics, have been given argumentation based characterisations [15,23,27]. Dung's theory thus provides a dialectical semantics for these logics, and the above-mentioned argument games can be viewed as alternative dialectical proof theories for these logics. The fact that reasoning in existing non-monotonic logics can thus be characterised, testifies to the generality of the dialectical principles of attack and reinstatement; principles that are also both intuitive and familiar in human modes of reasoning, debate and dialogue. Argumentation theory thus provides a characterisation of both human and logic-based reasoning in the presence of uncertainty and conflict, through the abstract dialectical modelling of the *process* whereby arguments can be moved to attack and reinstate/defend other arguments. The theory's value can therefore in large part be attributed to its explanatory potential for making non-monotonic reasoning processes inspectable and readily understandable for human users, and its underpinning of dialogical and more general communicative interactions that may involve heterogeneous (human and software) agents reasoning in the presence of uncertainty and conflict.

More recently, the *ASPIC* framework [18] was developed in response to the fact that the abstract nature of Dung's theory gives no guidance as to what kinds of instantiations satisfy intuitively rational properties. *ASPIC* was not designed from scratch but was meant to integrate, generalise and further develop existing work on structured argumentation, partly originating from before Dung's paper (e.g. [37,46,38,15,44]). *ASPIC* adopts an intermediate level of abstraction between Dung's fully abstract level and concrete instantiating logics, by making some minimal assumptions on the nature of the logical language and the inference rules, and then providing abstract accounts of the structure of arguments, the nature of attack, and the use of preferences. [18] then formulated consistency and closure postulates that cannot be formulated at the abstract level, and showed these postulates to hold for a special case of *ASPIC*; one in which preferences were *not* accounted for. In [40], *ASPIC*$^+$ then generalised *ASPIC* to accommodate a broader range of instantiations (including assumption-based argumentation [15] and systems using argument schemes), and showed that under some assumptions, the postulates were satisfied when applying preferences. [47] subsequently showed that the Carneades system [25] is an instance of *ASPIC*$^+$ with no defeat cycles.

In this paper we build on and modify [40]'s *ASPIC*$^+$ framework, to develop a more general structured framework for argumentation with preferences. We make three main contributions. We first motivate a revised definition of conflict free sets of arguments for *ASPIC*$^+$, adapt *ASPIC*$^+$ to accommodate a broader range of instantiating logics, and show that the resulting framework satisfies the key properties and postulates in [23] and [18]. Second, we formalise instantiation of the new framework by Tarskian (and in particular classical) logics extended with preferences, and demonstrate that such instantiations satisfy [18]'s rationality postulates. Third, we examine and counter recent critiques of Dung's framework and its extensions to accommodate preferences.[1]

With regard to the first contribution, Section 2 presents the conceptual foundations for our framework in the context of the above value proposition of argumentation as providing a bridging role between formal logic and human modes of reasoning. Specifically, we: (i) posit criteria for defining attack relations, given their dual role in declaratively denoting the mutual incompatibility of the information contained in the attacking arguments, and their dialectical use; (ii) motivate the distinction between preference dependent and preference independent attacks, where only the former's use in a dialectical context (as defeats) should be contingent upon preferences; (iii) argue that unlike current approaches [5,10,33], including [40]'s *ASPIC*$^+$, it is conceptually more intuitive to define conflict-free sets in terms of those that do not contain attacking arguments, so that defeats are only deployed dialectically. Section 3 then revisits and generalises [40]'s *ASPIC*$^+$ framework in light of Section 2's conceptual foundations. The new notion of conflict-free is adopted, and [40]'s *ASPIC*$^+$ framework is extended to accommodate instantiation by arguments with consistent premises, thus generalising the framework to accommodate a broader range of instantiations. Section 4 then presents key technical results. We show that Section 3's revised and generalised *ASPIC*$^+$ satisfies properties of Dung's theory and [18]'s rationality postulates.

Section 5 then presents the second main contribution, so testifying to the generality of the framework proposed here. To start with, we generalise results of [40], in which preferences defined over arguments on the basis of preorderings over arguments' constituent rules and premises, are shown to satisfy properties that ensure satisfaction of rationality postulates. In this paper we show that these properties are also satisfied by other ways of defining preferences, and furthermore address some limitations of [40]'s way of defining preferences. We then relate our work to Amgoud & Besnard's [2,3] recent 'abstract logic' approach to argumentation, which considers instantiations of Dung's framework by Tarskian logics. We combine this approach with the *ASPIC*$^+$ framework, and then extend [2,3]'s abstract logic approach with preferences, and also combine this extension with *ASPIC*$^+$. Given Section 4's results, these combinations imply that we are the first to show satisfaction of [18]'s rationality postulates for Tarskian logic instantiations with and without preferences. Following this, we reconstruct classical logic approaches to argumentation [12,13,26], including those that additionally accommodate preferences [5]. To the best of our knowledge, we are the first to prove [18]'s postulates for classical logic approaches with preferences. Finally, we show a correspondence between a particular classical logic instantiation of Section 3's *ASPIC*$^+$ framework and Brewka's preferred subtheories [16].

[1] The current paper extends [36] in which the revised definition of conflict free sets is first proposed, and *ASPIC*$^+$ is adapted to accommodate classical logic instantiations.

Section 6 discusses related work, and so presents our third main contribution. Specifically, we compare the generality of *ASPIC*$^+$ with the abstract logic proposal for structured argumentation, and argue that the latter only applies to deductive (e.g., classical logic) approaches, and not to mixed deductive and defeasible argumentation which requires modelling of defeasible inference rules. We also argue that inclusion of defeasible inference rules in models of argumentation is required if argumentation is to bridge the gap between formalisms and human reasoning, as defeasible reasons are an essential ingredient of human reasoning. Section 6 also counters a number of recent criticisms of Dung's abstract approach, as well as critiques of Dung's approach extended with preferences. We claim that a proper modelling of the use of preferences requires making the structure of arguments explicit.

2. Logic, argumentation and preferences

2.1. Background

A *Dung argumentation framework* (*AF*) [23] is a tuple $(\mathcal{A}, \mathcal{C})$, where $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation on the arguments $\mathcal{A}$. $S \subseteq \mathcal{A}$ is then said to be *conflict free* iff $\forall X, Y \in S$, $(X, Y) \notin \mathcal{C}$. The status of arguments is then evaluated as follows:

Definition 1. Let $(\mathcal{A}, \mathcal{C})$ be an *AF*. For any $X \in \mathcal{A}$, X is acceptable with respect to some $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $(Y, X) \in \mathcal{C}$ implies $\exists Z \in S$ s.t. $(Z, Y) \in \mathcal{C}$. Let $S \subseteq \mathcal{A}$ be *conflict free*. Then:

- S is an *admissible* extension iff $X \in S$ implies X is acceptable w.r.t. S;
- S is a *complete* extension iff $X \in S$ whenever X is acceptable w.r.t. S;
- S is a preferred extension iff it is a set inclusion maximal complete extension;
- S is the grounded extension iff it is the set inclusion minimal complete extension;
- S is a stable extension iff it is preferred and $\forall Y \notin S$, $\exists X \in S$ s.t. $(X, Y) \in \mathcal{C}$.

For $T \in$ {complete, preferred, grounded, stable}, X is *sceptically* or *credulously* justified under the T semantics if X belongs to all, respectively at least one, T extension.

A number of works [5,10,33] augment *AF*s to formalise the role of the relative strengths of arguments at the *abstract* level. The basic idea in all these works is that an attack by X on Y succeeds as a *defeat* only if Y is not stronger than X. For example, preference-based *AF*s (*PAF*s) [5] are tuples $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$, where given the preordering $\preccurlyeq \subseteq \mathcal{A} \times \mathcal{A}$, Y is stronger than X iff Y is strictly preferred to X ($X \prec Y$ iff $X \preccurlyeq Y$ and $Y \not\preccurlyeq X$). In [33], preferences between arguments are not based on a given preordering, but rather are themselves defeasible and possibly conflicting, and so are themselves the conclusions of arguments. In [40]'s *ASPIC*$^+$ framework, arguments are defined by strict and defeasible rules and premises expressed in some abstract language. Attacks between arguments are defined, and a preference relation over arguments is used to derive a defeat relation. Unlike *PAF*s and [10]'s value based *AF*s, *ASPIC*$^+$'s use of preferences to define defeat takes the structure of arguments into account.

In all the above approaches, the justified arguments are then evaluated on the basis of the derived defeat relation, rather than the original attack relation. In other words, a conflict free set is one that contains no two defeating arguments, and the defeat relation replaces the attack relation $\mathcal{C}$ in Definition 1.

Prior to discussing the role of, and relationship between attacks, preferences and defeats, recall that Section 1 discussed how abstract argumentation and argument game proof theories: (a) provide dialectical semantics, respectively proof theories, for non-monotonic reasoning, where; (b) the abstract modelling of the process whereby arguments are submitted to attack and defend, comports with intuitive human modes of reasoning and debate. Thus, the added value of argumentation is in large part due to its potential for facilitating dynamic, interactive and heterogeneous (both automated and human) reasoning in the presence of uncertain and conflicting knowledge.

It is in this context that we motivate criteria for defining attack relations, the role of preferences, and a new approach to defining the extensions of a framework in terms of both defeat *and* attack relations. In what follows we assume that arguments are built from strict (i.e., deductive) and defeasible inference rules (a distinction that is made more precise in Section 3 and further discussed in Section 6), and refer to an argument's *conclusion* following from its constituent *premises* and *rule* applications (referred to collectively as the argument's *support*).

2.2. The two roles of attacks

Attacks play two roles. Firstly, that X attacks Y, is an abstract, declarative representation of the mutual incompatibility of the information contained in the attacking arguments. Secondly, the attack abstractly characterises the dialectical use of X as a counter-argument to Y. The former role suggests a necessary condition for specifying an attack between X and Y, namely, that they contain mutually incompatible information. However the second role suggests that this condition is not sufficient; attacks should also be defined in such a way as to reflect their use in debate and discussion. Intuitively, if Y is proposed as an argument, then in seeking a counter-argument to Y, one seeks to construct an argument X whose conclusion is in conflict with the conclusion or some supporting element of Y. This motivates a definition of attack according to which

only an argument's *final* conclusion is relevant for whether it attacks another argument. For example, consider argument Y concluding *Tweety flies*, supported by the premise *Tweety is a bird* and the defeasible rule that *birds fly*. Consider also argument X concluding *Tweety does not fly*, supported by the premise and defeasible rule *Tweety is a penguin* and *penguins don't fly*. Then it is reasonable to say that X and Y attack each other, but if X is extended with the defeasible rule that *non-flying animals do not have wings*, to obtain X' claiming *Tweety does not have wings*, then X' should not attack Y, since its final conclusion does not conflict with any element of Y. Intuitively, X' would not be moved as a counter-argument to Y; rather it is the sub-argument X of X' that would be moved. An additional reason for not allowing X' to attack Y is that otherwise any continuation of X (and not just X') with further inferences would also attack Y, which may dramatically increase the number of attacks defined by a theory (and thus the computational expense incurred in evaluating the justified arguments). For example, if arguments can be constructed with the full power of classical logic, then this would yield an infinite number of attackers of Y.

A final requirement for attacks is that they should only be targeted at fallible elements of an argument, i.e., only on uncertain premises or defeasible inferences. In particular, conclusions of deductive inferences in an argument cannot be attacked. This should be obvious since the very meaning of deductive inference is that the truth of the premises of a deductive inference *guarantees* the truth of its conclusion. Any disagreement with the conclusion of a deductive inference should therefore be expressed as an attack on either uncertain premises or defeasible sub-arguments of the attacked argument. This informal analysis is supported by recent formal results [18,26] showing that allowing attacks on deductive inferences leads to violation of rationality postulates.

2.3. Distinguishing preference dependent and independent attacks

We now motivate the distinction between preference dependent and preference independent attacks. Firstly, note that we assumed above that arguments have three elements: a conclusion, a set of premises, and inference steps from the premises to the conclusion. Arguments can then in general be attacked in three ways: on their premises, on their conclusion and on their inference steps. We also argue that in practice, preferences are often used in argumentation, so that a formal framework that aims to bridge the gap with human modes of argumentation, should accommodate preferences as first class citizens, instead of implicitly encoding them by other means (such as with explicit exception or applicability predicates). We now discuss to what extent these three types of attacks require preferences to succeed as defeats. To start with, we claim that attacks on conclusions should be resolved with preferences, since such attacks arise because of conflicting reasons for and against a conclusion. In such cases, explicit preferences are used to resolve such conflicts, e.g., based on rule priorities in legal systems, orderings on desires or values in practical reasoning, or reliability orderings in epistemic reasoning. For example, consider the above symmetrically attacking arguments X and Y respectively concluding *Tweety does not fly* and *Tweety flies*. Based on the specificity principle's prioritisation of properties of sub-classes over super-classes, one preferentially concludes *Tweety does not fly*. The use of the specificity principle can be modelled at the meta-level (i.e., meta to the object-level logic in which arguments X and Y are constructed), as a preference for X over Y, so that X asymmetrically defeats Y.

However, assuming sufficient expressive power, one could also encode this meta-level arbitration of the conflict in the object level logic, as undercutting attacks on inference steps [37]. The inferential step licensed by the rule $bf = \textit{birds fly}$, is blocked by a rule pNf that states that if the bird is a penguin, then the inferential step encoded in the rule bf, is not valid. This suggests the use of undercut attacks on an inference step for yielding the same results as those obtained through the use of preferences, in a way that makes the rationale for preference application more explicit. Undercuts also yield effects that cannot be exclusively effected through preferences. Consider Pollock's classic example [37] in which *there is a red light shining* undercuts the rule *if an object looks red then it is red*, so blocking the inference from *there is an object that looks red*, to the conclusion *the object is red*. Here, the undercut effectively expresses a preference for not drawing the inference over drawing the inference; something that cannot be expressed as a preference ordering over arguments.

We conclude that when specifying an attack by Z on Y, based on Z's conclusion undercutting a rule in Y, the attacking argument is first and foremost expressing reasons for preferring not to infer Y's conclusion over inferring Y's conclusion. Such attacks should therefore be 'preference independent', since qualifying the success of such an attack (as a defeat) as being contingent on Y not being preferred to Z, would be to contradict the preference that is effectively expressed by the attack itself. In other words, a priority relation that regards the undercut rule as of higher priority than the undercutting rule cannot be regarded as a preference for drawing the inference over not drawing the inference, since the opposite preference (for not drawing, over drawing, the inference) is already expressed in the undercutter. Thus, we argue for a distinction between *preference dependent* and *preference independent* attacks, where undercuts fall into the latter category. Note that this does not preclude that a third argument Z' attacks Z's conclusion that Y's conclusion should not be inferred, where Z''s attack is preference dependent.

Finally, we claim that whether attacks on premises are preference-dependent, depends on the nature of the premise that is attacked. Normally, preferences are needed except if the premise states some kind assumption in the absence of evidence to the contrary, as, for example, negation as failure assumptions in logic programming. If Y makes use of a negation as failure assumption of the form $\sim\alpha$, denoting that 'α is not provable', then an argument Z concluding α, preference independent attacks Y, since the construction of Z is contingent on the non-provability of α, i.e., the absence of an acceptable argument Y concluding α.

2.4. The distinct uses of attacks and defeats

To recap, attacks encode the mutual incompatibility of the information contained in the attacking and attacked arguments, in a way that accounts for their dialectical use. In turn, the dialectical use of attacks as defeats may or may not be contingent on the preferences defined over the arguments.

As described in Section 2.1, existing works that account for preferences and/or values [5,10,33], including [40]'s *ASPIC*$^+$ framework, define conflict-free and acceptable sets of arguments with respect to the defeats. However, we argue that defining conflict free sets in terms of defeats is conceptually wrong. Since attacks indicate the mutual incompatibility of the information contained in the attacking and attacked arguments, then intuitively one should continue to define conflict-free sets in terms of those that do not contain attacking arguments. Defeats only encode the preference dependent use of attacks in the dialectical evaluation of the acceptability of arguments. They have no bearing on whether one argument can be said to be logically incompatible with another, but rather whether the attack can be validly employed in a dialectical setting.

In the following section, we therefore re-define [40]'s *ASPIC*$^+$ notion of a conflict free set, as one in which no two arguments *attack* rather than defeat. We then examine the implications of this in Section 4.2.

3. The ASPIC$^+$ framework

In this section we review [40]'s *ASPIC*$^+$ framework in light of the criteria and requirements enumerated in Sections 2.2 and 2.3. We also modify the framework in two ways: 1) we change the definition of conflict free, as proposed above; 2) we further generalise *ASPIC*$^+$ so as to capture deductive approaches to argumentation [2,3,5,13]. In addition, we simplify some of [40]'s notations and definitions.

3.1. ASPIC$^+$ arguments

The *ASPIC*$^+$ framework defines arguments, as in [48], as inference trees formed by applying strict or defeasible inference rules to premises that are well-formed formulae (wff) in some logical language. The distinction between two kinds of inference rules is taken from [37,30,39,48]. Informally, if an inference rule's antecedents are accepted, then if the rule is strict, its consequent must be accepted *no matter what*, while if the rule is defeasible, its consequent must be accepted *if there are no good reasons not to accept it*. Arguments can be attacked on their (non-axiom) premises and on their applications of defeasible inference rules. Some attacks succeed as *defeats*, which is partly determined by preferences. The acceptability status of arguments is then defined by applying any of [23]'s semantics for abstract argumentation frameworks to the resulting set of arguments with its defeat relation.

We emphasise that *ASPIC*$^+$ is not a system but a framework for specifying systems. It defines the notion of an abstract *argumentation system* (a notion adapted from [48]) as a structure consisting of a logical language $\mathcal{L}$ with a binary relation $^-$, a naming convention n for defeasible rules and a set $\mathcal{R}$ consisting of two subsets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules. (As is usual, inference rules are defined over the language $\mathcal{L}$, and are not elements in the language.) *ASPIC*$^+$ as a framework does not make any assumptions on how these elements are defined in a given argumentation system (the idea to abstract from the precise nature of $\mathcal{L}/\mathcal{R}$ is taken from [30,48,15] while the idea to abstract from $^-$ and n is taken from [15] and [39], respectively).

ASPIC$^+$'s inference rules can be used in two ways: they could encode domain-specific information but they could also express general laws of reasoning. When used in the latter way, the defeasible rules could, for example, express argument schemes [49], while the strict rules could be determined by the choice of the logical language $\mathcal{L}$: its formal semantics will then tell which inference rules over $\mathcal{L}$ are valid and can therefore be added to $\mathcal{R}_s$. If the strict rules are thus chosen then they could consist, for example, of all classically valid inferences or more generally conform to any Tarskian consequence notion (cf. [2]). Notice that inclusion of defeasible rules in *ASPIC*$^+$ requires some explanation, given that much current work formalises construction of arguments as deductive [2,3], and in particular classical [13,26] inference. We justify the need for inclusion of defeasible inference rules in Section 6.

As just explained, the basic notion of *ASPIC*$^+$ is that of an argumentation system. Arguments are then constructed with respect to a knowledge base. Definitions of these are taken from [40] (with some modifications that will be subsequently described).

Definition 2 *(ASPIC*$^+$ *argumentation system).* An *argumentation system* is a tuple $AS = (\mathcal{L}, ^-, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language.
- $^-$ is a function from $\mathcal{L}$ to $2^{\mathcal{L}}$, such that:
  - φ is a *contrary* of ψ if $\varphi \in \overline{\psi}$, $\psi \notin \overline{\varphi}$;
  - φ is a *contradictory* of ψ (denoted by '$\varphi = -\psi$'), if $\varphi \in \overline{\psi}$, $\psi \in \overline{\varphi}$;
  - each $\varphi \in \mathcal{L}$ has at least one contradictory.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over wff in $\mathcal{L}$), and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $n : \mathcal{R}_d \rightarrow \mathcal{L}$ is a naming convention for defeasible rules.

Intuitively, contraries can be used to model well-known constructs like negation as failure in logic programming or consistency checks in default logic. Note that we illustrate requirements for the asymmetric notion of contrary (as opposed to the more familiar symmetric notion of contradictory associated standardly with negation) in Section 3.2. Note also that in previous publications on *ASPIC*$^+$ (including [40]) the idea of a naming convention n was instead informally introduced when defining undercutting attack (see Definition 8 below). Informally, $n(r)$ is a wff in $\mathcal{L}$ which says that the defeasible rule $r \in \mathcal{R}$ is applicable.

Definition 3. For any $S \subseteq \mathcal{L}$, let the *closure of S under strict rules*, denoted $Cl_{R_s}(S)$, be the smallest set containing S and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in $Cl_{R_s}(S)$. Then a set $S \subseteq \mathcal{L}$ is

- *directly consistent* iff $\nexists \psi$, $\varphi \in S$ such that $\psi \in \overline{\varphi}$;
- *indirectly consistent* iff $Cl_{R_s}(S)$ is directly consistent.

This definition is generalised from [18], in which these two notions of consistency were defined for the special case where $^-$ corresponds to negation.

Definition 4 *(ASPIC$^+$ knowledge base).* A *knowledge base* in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the *axioms*) and $\mathcal{K}_p$ (the *ordinary premises*).

Intuitively, the axioms are certain knowledge and thus cannot be attacked, whereas the ordinary premises are uncertain and thus can be attacked. The distinction between ordinary premises and axiom premises is needed to capture systems like, for instance, Pollock's system [37], which does not allow attacks on premises, and which therefore need to be modelled as axiom premises. In [40], the knowledge base was also assumed to have *issue* and *assumption* premises, which were used, respectively, to prove that Carneades [25] and assumption-based argumentation [15] are special cases of *ASPIC*$^+$. In the present paper we omit issue premises for simplicity while, as further discussed below in Section 6.1, [40]'s result on assumption-based argumentation also holds if all premises are ordinary instead of assumption premises. Furthermore, in previous *ASPIC*$^+$ publications (including [40]) we included preorderings on $\mathcal{R}_d$ and $\mathcal{K}_p$ in the definitions of argumentation systems and knowledge bases respectively. We remove references to these preorderings in the above general definitions, and only introduce them when they are required for defining preference orderings over arguments.

Example 1. Let $(\mathcal{L}, ^-, \mathcal{R}, n)$ be an argumentation system where:

- $\mathcal{L}$ is a language of propositional literals, composed from a set of propositional atoms $\{a, b, c, \ldots\}$ and the symbols $\neg$ and $\sim$ respectively denoting strong and weak negation (i.e., negation as failure). α is a strong literal if α is a propositional atom or of the form $\neg\beta$ where β is a propositional atom (strong negation cannot be nested). α is a wff of $\mathcal{L}$, if α is a strong literal or of the form $\sim\beta$ where β is a strong literal (weak negation cannot be nested).
- $\alpha \in \overline{\beta}$ iff (1) α is of the form $\neg\beta$ or β is of the form $\neg\alpha$; or (2) β is of the form $\sim\alpha$ (i.e., for any wff α, α and $\neg\alpha$ are contradictories and α is a contrary of $\sim\alpha$).
- $\mathcal{R}_s = \{t, q \rightarrow \neg p\}$, $\mathcal{R}_d = \{\sim s \Rightarrow t; r \Rightarrow q; a \Rightarrow p\}$.
- $n(\sim s \Rightarrow t) = d_1$, $n(r \Rightarrow q) = d_2$, $n(a \Rightarrow p) = d_3$.

Furthermore, $\mathcal{K}$ is the knowledge base such that $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{a, r, \neg r, \sim s\}$.

Arguments are defined below (as in [40]), together with some associated notions. Informally, for any argument A, `Prem` returns all the formulae of $\mathcal{K}$ (*premises*) used to build A, `Conc` returns A's conclusion, `Sub` returns all of A's sub-arguments, `DefRules` and `StRules` respectively return all defeasible and all strict rules in A, and `TopRule`(A) returns the last rule applied in A.

Definition 5 *(ASPIC$^+$ arguments).* An *argument* A on the basis of a knowledge base $\mathcal{K}$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n)$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{Rules}(A) = \emptyset$; $\texttt{TopRule}(A) =$ undefined.
2. $A_1, \ldots, A_n \rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$ in $\mathcal{R}_s$.
   $A_1, \ldots, A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$.
   $\texttt{Prem}(A)$[2] $= \texttt{Prem}(A_1) \cup \cdots \cup \texttt{Prem}(A_n)$,

[2] Note that all premises in ASPIC$^+$ arguments are used in deriving its conclusion, so enforcing a notion of relevance analogous to the subset minimality condition requirement on premises in classical logic approaches to argumentation (see Section 5.2).

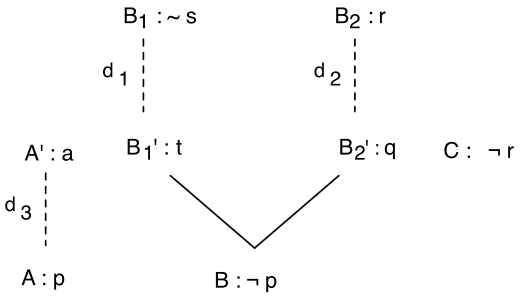

Fig. 1. *ASPIC*$^+$ arguments and their conclusions, with dashed and solid lines respectively representing application of defeasible and strict inference rules.

$\texttt{Conc}(A) = \psi$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \cdots \cup \texttt{Sub}(A_n) \cup \{A\}$. Note that $A_1 \ldots A_n$ are referred to as the *proper* sub-arguments of A.
$\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \cdots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow / \Rightarrow \psi\}$.
$\texttt{DefRules}(A) = \{r | r \in \texttt{Rules}(A), r \in \mathcal{R}_d\}$.
$\texttt{StRules}(A) = \{r | r \in \texttt{Rules}(A), r \in \mathcal{R}_s\}$.
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow / \Rightarrow \psi$.

Furthermore, for any argument A:

- $\texttt{Prem}_n(A) = \texttt{Prem}(A) \cap \mathcal{K}_n$ and $\texttt{Prem}_p(A) = \texttt{Prem}(A) \cap \mathcal{K}_p$.
- If $\texttt{DefRules}(A) = \emptyset$, then $\texttt{LastDefRules}(A) = \emptyset$, else;
  if $A = A_1, \ldots, A_n \Rightarrow \phi$ then $\texttt{LastDefRules}(A) = \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \phi\}$, otherwise $\texttt{LastDefRules}(A) = \texttt{LastDefRules}(A_1) \cup \cdots \cup \texttt{LastDefRules}(A_n)$.
- A is: *strict* if $\texttt{DefRules}(A) = \emptyset$; *defeasible* if $\texttt{DefRules}(A) \neq \emptyset$; *firm* if $\texttt{Prem}(A) \subseteq \mathcal{K}_n$; *plausible* if $\texttt{Prem}(A) \not\subseteq \mathcal{K}_n$; *fallible* if A is plausible or defeasible; *finite* if $\texttt{Rules}(A)$ is finite.[3]

Henceforth, we may employ the following notation for arguments.

Notation 2 *(Notation for arguments).*

1. $S \vdash \varphi$ may be written to denote that there exists a strict argument A such that $\texttt{Conc}(A) = \varphi$, with all premises taken from S (i.e., $\texttt{Prem}(A) \subseteq S$).
2. Arguments may be written as lists of premises and rules separated by semi-colons, or in the case that an argument has a top rule, we may write such an argument as the top rule with the antecedents replaced by the names of the sub-arguments that conclude the antecedents. For example, we may write $A = [s; s \Rightarrow r; q; r, q \rightarrow \neg p]$ or $A = [A_1, A_2 \rightarrow \neg p]$, where $A_1 = [s; s \Rightarrow r]$, $A_2 = [q]$.
3. Letting Γ be a set of arguments, we may as an abuse of notation write $\texttt{F}(\Gamma)$ to denote $\bigcup_{A \in \Gamma} \texttt{F}(A)$, where $\texttt{F} \in \{\texttt{Prem}, \texttt{Conc}, \texttt{Sub}, \texttt{Rules}, \texttt{TopRule}, \texttt{DefRules}, \texttt{StRules}\}$.

Example 3. The arguments (shown in Fig. 1) defined on the basis of the knowledge base and argumentation system in Example 1 are: $A' = [a]$, $A = [A' \Rightarrow p]$, $B_1 = [\sim s]$, $B'_1 = [B_1 \Rightarrow t]$, $B_2 = [r]$, $B'_2 = [B_2 \Rightarrow q]$, $B = [B'_1, B'_2 \rightarrow \neg p]$, $C = [\neg r]$.

Furthermore, $\texttt{Prem}(B) = \{\sim s, r\}$; $\texttt{Conc}(B) = \neg p$; $\texttt{Sub}(B) = \{B_1, B_2, B'_1, B'_2\}$; $\texttt{TopRule}(B) = t, q \rightarrow \neg p$; $\texttt{DefRules}(B) = \{\sim s \Rightarrow t, r \Rightarrow q\}$; $\texttt{StRules}(B) = \{t, q \rightarrow \neg p\}$.

We now adapt [40]'s above definition of an argument so as to consider a special class of arguments whose premises are 'c-consistent' (for "contradictory-consistent"). We thus generalise *ASPIC*$^+$ so as to accommodate deductive approaches to argumentation [2,3,5,13] that require that the arguments defined by the instantiating logic have consistent premises.

Definition 6 *(c-consistent).* A set $S \subseteq \mathcal{L}$ is *c-consistent* if for no φ it holds that $S \vdash \varphi, -\varphi$. Otherwise S is *c-inconsistent*. We say that $S \subseteq \mathcal{L}$ is minimally c-inconsistent iff S is c-inconsistent and $\forall S' \subset S$, S' is c-consistent.

Note that we use the term 'c-consistent' to distinguish the notion of consistency in Definition 2. Also note that if $S \vdash \varphi, \phi$, where $\phi \in \overline{\varphi}$, then S can still be c-consistent. As we will see later, such situations do not arise when capturing deductive approaches in *ASPIC*$^+$, as in these approaches there are no contraries, only contradictories.

[3] As explained in [40], Definition 5 allows for arguments that are 'backwards' infinite in that they do not 'bottom' out in premises from the knowledge base.

Definition 7 *(c-consistent argument).* An *argument* A on the basis of a knowledge base $\mathcal{K}$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n)$, is c-consistent iff $\texttt{Prem}(A)$ is c-consistent.

3.2. Attacks and defeats

We now review [40]'s definition of attacks and defeats amongst arguments. An argument A attacks an argument A' if the conclusion of A (i.e., $\texttt{Conc}(A)$) is a contrary or contradictory of: an ordinary premise in A'; the consequent of a defeasible rule in A', or; a defeasible inference step in A'. These three kinds of attack are respectively called undermining, rebutting and undercutting attacks.

Definition 8 *(ASPIC$^+$ attacks). A attacks B* iff *A undercuts*, *rebuts* or *undermines B*, where:

- *A undercuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{n(r)}$ for some $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.
- *A rebuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' \in \texttt{Sub}(B)$ of the form $B'_1, \ldots, B'_n \Rightarrow \varphi$. In such a case *A contrary-rebuts B* iff $\texttt{Conc}(A)$ is a contrary of φ.
- Argument *A undermines B* (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' = \varphi$, $\varphi \in \texttt{Prem}_p(B)$. In such a case *A contrary-undermines B* iff $\texttt{Conc}(A)$ is a contrary of φ.

Example 4 *(Example 3 continued).* For the arguments in Example 3, B rebuts A on A, C undermines B and B'_2 on B_2, and C and B_2 undermine each other. The attack graph is shown in Fig. 2(a). Notice that if in addition one had the argument $D = [\sim d; \sim d \Rightarrow s]$, then D would contrary-undermine B and B'_1 on B_1. Moreover, if in addition one had the argument $E = [r; r \rightarrow \neg d_3]$, then E would undercut A on A.

Note that Definition 8 complies with Section 2.2 and 2.3's rationale for defining attacks. An attack originating from an argument A requires that its conclusion $\texttt{Conc}(A)$ (and not the conclusion of any sub-argument of A) be in conflict with some *fallible* element – i.e., some ordinary premise, or defeasible rule or conclusion of a defeasible rule – in the attacked argument. Thus, while B_2 rebut-attacks C in Example 4, the argument B, that contains B_2 as a sub-argument, does not attack C. Also, although A and B have contradictory conclusions, only B rebuts A; A does not rebut B as B's conclusion is the consequent of a strict rule. [18] refers to this as a *restricted rebut*, and shows for a special case of *ASPIC*$^+$ that if the restriction is lifted so as to allow A to rebut B, then this could lead to violation of [18]'s rationality postulates.

Attacks can then be distinguished as to whether they are preference-dependent or preference-independent, where the former's success as defeats is determined by a preference ordering $\preccurlyeq$ on the constructed arguments. We make no assumptions on the properties of $\preccurlyeq$. In Section 5.1 we will utilise two preorderings $\leqslant$ on defeasible rules and $\leqslant'$ on ordinary premises to give example definitions of $\preccurlyeq$, but the definition of defeat does not rely on these preorderings. As usual:

- the strict counterpart $\prec$ of $\preccurlyeq$ is defined as $X \prec Y$ iff $X \preccurlyeq Y$ and $Y \not\preccurlyeq X$, and;
- $X \approx Y$ denotes that $X \preccurlyeq Y, Y \preccurlyeq X$.

Definition 9 *(ASPIC$^+$ defeats).* Let A attack B on B'. If A undercut, contrary-rebut, or contrary-undermine attacks B on B' then A is said to *preference-independent* attack B on B', otherwise A is said to *preference-dependent* attack B on B'.

Then, A **defeats** B iff for some B' either A *preference-independent* attacks B on B', or A *preference-dependent* attacks B on B' and $A \not\prec B'$.

A strictly *defeats* B iff A *defeats* B and B does not *defeat* A.

Notation 5. Henceforth, $\rightharpoonup$ may denote the attack relation, and $\hookrightarrow$ the defeat relation.

The definition of defeats complies with Section 2.3's rationale for distinguishing between preference dependent and preference independent attacks. Undercuts always succeed as defeats, and so are preference independent. As discussed in Section 2.3, undercutting attacks encode an asymmetry: the use of the attacked rule named r in an argument B is contingent on the absence of an acceptable attacking argument A with an undercutting, conclusion $\neg r$. The notion of a contrary relation generalises the above cases of asymmetric preference independent attacks, providing for greater flexibility in declaring formulae φ and ψ incompatible, where attacks from φ to ψ are not undercuts but are still preference independent.[4] As discussed in Section 2.3 an example of such a cases is when ψ is a negation as failure assumption. This is illustrated in Example 1 in which α is a contrary of $\sim\alpha$, so that an undermining attack from an argument A concluding α on an ordinary premise $\sim\alpha$ in an argument B, is preference independent.

[4] Notice that in such cases it would be counter-intuitive to allow ψ to be an axiom premise or the conclusion of a strict rule. In the rest of this paper we will therefore assume that such cases do not arise. This assumption is formalised in Definition 12.

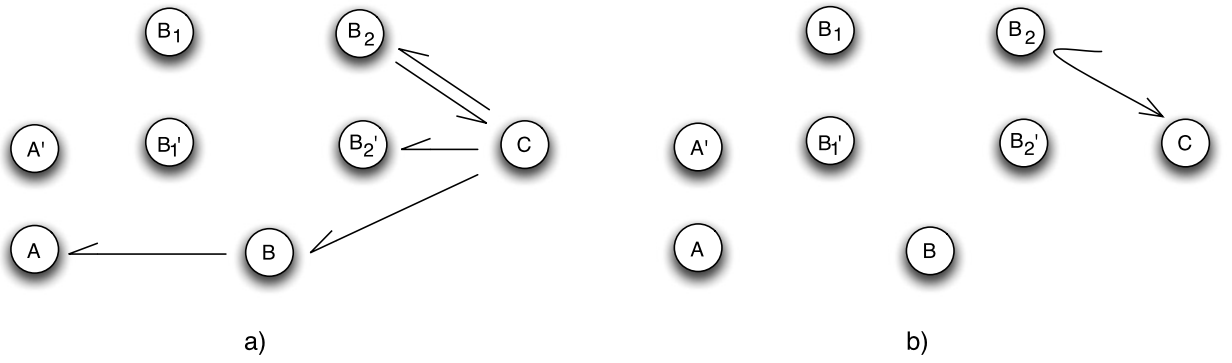

Fig. 2. Example 4's *ASPIC*$^+$ attack graph shown in (a). Note that the dashed and solid lines representing application of defeasible and strict inference rules have been removed. The defeat graph (see Example 6) is shown in (b).

Example 6 *(Example 4 continued).* We assume that the argument ordering $\preccurlyeq$ is defined in terms of preorderings $\leqslant$ on defeasible rules and $\leqslant'$ on ordinary premises (in ways fully specified in Section 5.1 below). Assume that $r \Rightarrow q < a \Rightarrow p$ and $\neg r <' r$; $\neg a \approx' r$; $\sim s <' \neg r$. (As usual, $l \approx l'$ iff $l \leqslant l'$ and $l' \leqslant l$, while $l < l'$ iff $l \leqslant l'$ and $l' \not\leqslant l$; likewise for $\approx'$ and $<'$.)

Now let $B'_2 \prec A$, $B \prec A$ (because of $r \Rightarrow q < a \Rightarrow p$), $C \prec B_2$, $C \prec B'_2$, $C \prec B$ (because of $\neg r <' r$). Then B does not defeat A ($B \not\mapsto A$), $C \not\mapsto B$, $C \not\mapsto B'_2$ and $B_2 \hookrightarrow C$ (the arguments and defeats are depicted in Fig. 2(b)).

Note that if one had the additional arguments D and/or E described in Example 4, then D would defeat B_1 and so B'_1 and B, while E would defeat A. Note that $D \rightharpoonup B_1$ is preference independent since s is a contrary of $\sim s$; the validity of B_1, B'_1 and B is contingent on s not being provable (i.e., there being no acceptable argument for s).

3.3. Structuring argumentation frameworks

We now define two notions of a structured argumentation framework instantiated by an argumentation theory. The first is defined as in [40]. The second accounts for this paper's definition of c-consistent arguments.

Definition 10 *(Argumentation theory).* An *argumentation theory* is a tuple $AT = (AS, \mathcal{K})$ where AS is an argumentation system and $\mathcal{K}$ is a knowledge base in AS.

Definition 11 *((c-)structured argumentation frameworks).* Let AT be an argumentation theory $(AS, \mathcal{K})$.

- A *structured argumentation framework (SAF)* defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preccurlyeq \rangle$ where $\mathcal{A}$ is the set of all finite arguments constructed from $\mathcal{K}$ in AS (henceforth called the set of arguments on the basis of AT), $\preccurlyeq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.
- A *c-structured argumentation framework (c-SAF)* defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preccurlyeq \rangle$ where $\mathcal{A}$ is the set of all c-consistent finite arguments constructed from $\mathcal{K}$ in AS, $\preccurlyeq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

Henceforth, we may write '*(c-)SAF*' instead of writing '*SAF* or *c-SAF*'. Note that a c-SAF is a SAF in which all arguments are required to have a c-consistent set of premises.

In [40], it is assumed that any argumentation theory satisfies a number of properties. We repeat these here, and add an additional 'c-classicality' property for c-*SAF*s, in which we refer to the notion of 'closure under strict rules' and the notation '$S \vdash \phi$' given in Definition 3 and Notation 2 respectively.

Definition 12 *(Well defined (c-)SAFs).* Let $AT = (AS, \mathcal{K})$ be an argumentation theory, where $AS = (\mathcal{L}, ^-, \mathcal{R}, n)$. We say that AT is:

- *closed under contraposition* iff for all $S \subseteq \mathcal{L}$, $s \in S$ and ϕ, if $S \vdash \phi$, then $S \backslash \{s\} \cup \{-\phi\} \vdash -s$.
- *closed under transposition*[5] iff if $\phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{R}_s$, then for $i = 1 \ldots n$, $\phi_1, \phi_{i-1}, -\psi, \phi_{i+1}, \ldots, \phi_n \rightarrow -\phi_i \in \mathcal{R}_s$;
- *axiom consistent* iff $Cl_{\mathcal{R}_s}(\mathcal{K}_n)$ is consistent.
- *c-classical* iff for any minimal c-inconsistent $S \subseteq \mathcal{L}$ and for any $\varphi \in S$, it holds that $S \setminus \{\varphi\} \vdash -\varphi$ (i.e., amongst all arguments defined there exists a strict argument with conclusion $-\varphi$ with all premises taken from $S \setminus \{\varphi\}$).
- *well formed* if whenever φ is a contrary of ψ then $\psi \notin \mathcal{K}_n$ and ψ is not the consequent of a strict rule.[6]

[5] The notion of closure under transposition is taken from [18].

[6] This formulation repairs an error in the one of [40], which allowed for counterexamples to some results.

If a *c-SAF* is defined by an AT that is c-classical, axiom consistent, well formed and closed under contraposition or closed under transposition, then the *c-SAF* is said to be *well defined*.

If a *SAF* is defined by an AT that is axiom consistent, well formed and closed under contraposition or closed under transposition, then the *SAF* is said to be *well defined*.

Henceforth, we will assume that any *(c-)SAF* is well defined. The intuitions underlying the first four properties are self-evident. The rationale for the well-formed assumption is discussed in Section 3.2.

*(c-)SAF*s can now be linked to Dung frameworks. Firstly, note that as with existing approaches [5,10,33], [40]'s notion of a conflict free set of arguments is defined with respect to the derived defeat relation.

Definition 13 *(Defeat conflict free for (c-)SAFs).* Let $\Delta = \langle \mathcal{A}, \mathcal{C}, \preccurlyeq \rangle$ be a *(c-)SAF*, and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$, where $(X, Y) \in \mathcal{D}$ iff X defeats Y according to Definition 9. Then $S \subseteq \mathcal{A}$ is *defeat conflict free* iff $\forall X, Y \in S$, $(X, Y) \notin \mathcal{D}$.

However, we have in Section 2.4 argued that conflict free sets should be defined with respect to the attack relation, and defeats reserved for the dialectical use of attacks:

Definition 14 *(Attack conflict free for (c-)SAFs).* Let $\Delta = \langle \mathcal{A}, \mathcal{C}, \preccurlyeq \rangle$ be a *(c-)SAF*. Then $S \subseteq \mathcal{A}$ is *attack conflict free* iff $\forall X, Y \in S$, $(X, Y) \notin \mathcal{C}$.

In either case, the justified arguments are then evaluated on the basis of the extensions of a Dung framework instantiated by the arguments and derived defeat relation:

Definition 15 *(Extensions and justified arguments/conclusions of (c-)SAFs).* Let $\Delta = \langle \mathcal{A}, \mathcal{C}, \preccurlyeq \rangle$ be a *(c-)SAF*, and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$, where $(X, Y) \in \mathcal{D}$ iff X defeats Y. Let $S \subseteq \mathcal{A}$ be defeat or attack conflict free. The *extensions and justified arguments of* Δ are the extensions of the Dung framework $(\mathcal{A}, \mathcal{D})$, as defined in Definition 1.

For $T \in$ {admissible, complete, preferred, grounded, stable}, we say that:

- φ is a T credulously justified conclusion of Δ iff there exists an argument A. such that $\mathtt{Conc}(A) = \varphi$, and A is credulously justified under the T semantics.
- φ is a T sceptically justified conclusion of Δ iff for every T extension E, there exists an argument $A \in E$ such that $\mathtt{Conc}(A) = \varphi$.

S is a *def-T* extension if S is defined as defeat conflict free, and an *att-T* extension if S is defined as attack conflict free.

We now recall a definition from [40], and then in this paper we define the notion of an argument A being a strict continuation of a set of arguments $\{A_1, \ldots, A_n\}$.

Definition 16. The set $M(B)$ of the *maximal fallible sub-arguments* of B is defined such that for any $B' \in \mathtt{Sub}(B)$, $B' \in M(B)$ iff:

1. B''s top rule is defeasible or B' is an ordinary premise, and;
2. there is no $B'' \in \mathtt{Sub}(B)$ s.t. $B'' \neq B$ and $B' \in \mathtt{Sub}(B'')$, and B'' satisfies 1).

The maximal fallible sub-arguments of an argument B are those with the 'last' defeasible inferences in B or else (if B is strict) they are B's ordinary premises. That is, they are the maximal sub-arguments of B on which B can be attacked. In Example 3 we have that $M(A) = \{A\}$, $M(B) = \{B'_1, B'_2\}$, $M(C) = \{C\}$.

Definition 17 *(Strict continuations of arguments).* For any set of arguments $\{A_1, \ldots, A_n\}$, the argument A is a *strict continuation* of $\{A_1, \ldots, A_n\}$ iff:

- $\mathtt{Prem}_p(A) = \bigcup_{i=1}^{n} \mathtt{Prem}_p(A_i)$
  (i.e., the ordinary premises in A are exactly those in $\{A_1, \ldots, A_n\}$);
- $\mathtt{DefRules}(A) = \bigcup_{i=1}^{n} \mathtt{DefRules}(A_i)$
  (i.e., the defeasible rules in A are exactly those in $\{A_1, \ldots, A_n\}$);
- $\mathtt{StRules}(A) \supseteq \bigcup_{i=1}^{n} \mathtt{StRules}(A_i)$ and $\mathtt{Prem}_n(A) \supseteq \bigcup_{i=1}^{n} \mathtt{Prem}_n(A_i)$
  (i.e., the strict rules and axiom premises of A are a superset of the strict rules and axiom premises in $\{A_1, \ldots, A_n\}$).

Example 7 *(Example 3 continued).* In Example 3, we have that B is a strict continuation of B'_1 and B'_2. Now notice that the argumentation theory in Example 1 is not well defined, since it is neither closed under contraposition or transposition.

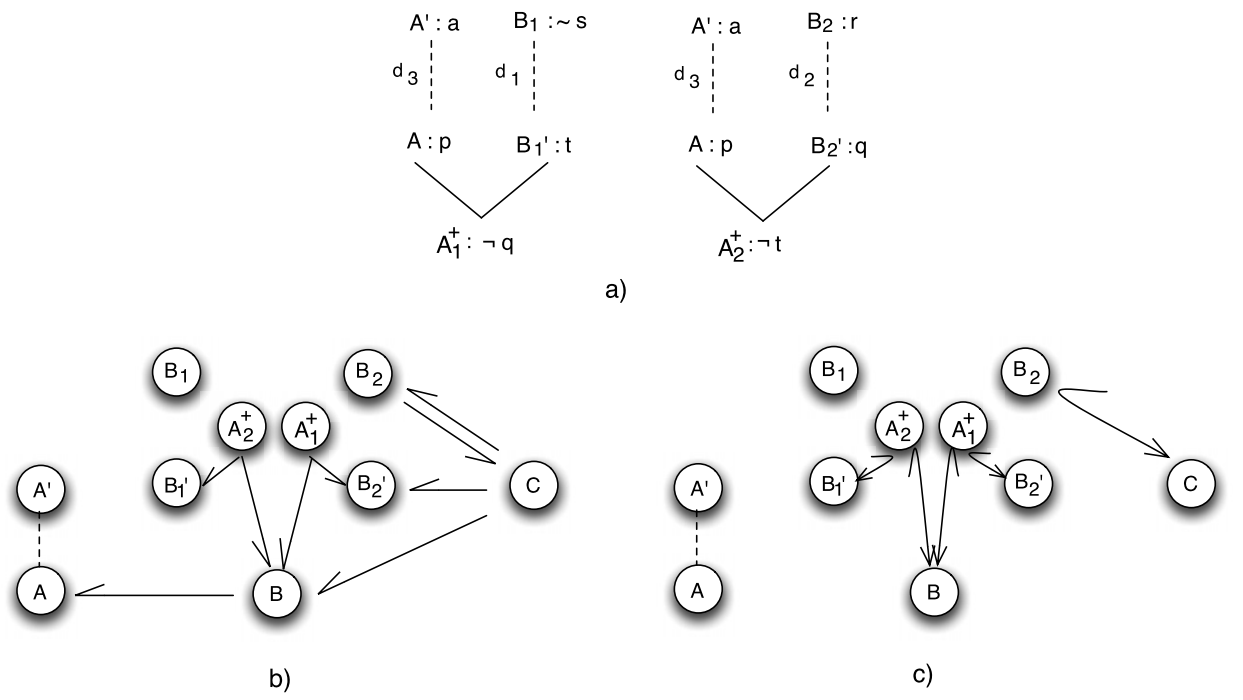

Fig. 3. Example 7's arguments A_1^+ and A_2^+ built from transpositions of the strict rule $t, q \rightarrow \neg p$ are shown in (a). *ASPIC*$^+$ attack graph shown in (b). A possible defeat graph is shown in (c).

Closure under transposition augments $\mathcal{R}_s$ with rules $t, p \rightarrow \neg q$ and $p, q \rightarrow \neg t$, so obtaining the additional arguments $A_1^+ = [B'_1, A \Rightarrow \neg q]$ that rebut-attacks B'_2 (and so B), and $A_2^+ = [A, B'_2 \Rightarrow \neg t]$ that rebut-attacks B'_1 (and so B). Fig. 3(a) shows these additional arguments and attacks.

4. Properties and postulates

In this section we examine the implications of the attack definition of conflict free sets. We show that under some assumptions on the preference ordering over arguments, both *SAF*s and *c-SAF*s satisfy the key properties of Dung frameworks. We also show that [18]'s rationality postulates straightforwardly hold. On the other hand, we will show that under [40]'s 'defeat definition' of conflict free, key properties of Dung frameworks straightforwardly hold, whereas satisfaction of [18]'s rationality postulates requires assumptions on preference orderings. Note that for the defeat definition, [40] has already shown satisfaction of the rationality postulates for *SAF*s. This paper extends [40]'s results to *c-SAF*s. Finally, we will show equivalence of admissible and complete extensions under the attack and defeat definitions of conflict free.

4.1. Properties of SAFs and c-SAFs under the attack definition of conflict free

Defining conflict free sets in terms of the attack relation, while using the defeat relation for determining the acceptability of arguments, potentially undermines some key results shown for Dung frameworks. To illustrate, consider Example 6's *SAF*, with the arguments and attacks shown in Fig. 2(a). As shown in Fig. 2(b), no argument defeats B, so $\{B\}$ is *att*-admissible (as defined in Definition 15). Since $B \prec A$, then $B \not\rightarrow A$, and so A is acceptable w.r.t. $\{B\}$. But $\{A, B\}$ is not attack conflict free and so not *att*-admissible. This violates Dung's *fundamental lemma* [23], which states that if S is admissible and A is acceptable w.r.t. S then $S \cup \{A\}$ is admissible. However, if the *SAF* is well defined (Definition 12), then under the assumption that an argument ordering is *reasonable*, we can show that the fundamental lemma holds.

An argument ordering $\preccurlyeq$ is reasonable if it satisfies properties that one might expect to hold of orderings over arguments composed from fallible and infallible elements. Firstly, whenever an argument A is not fallible (i.e., strict and firm), then it is strictly preferred over all arguments with fallible elements, and not less preferred than any other argument. Also, continuing an argument with only axiom premises and strict inferences does not change its relative preference. The second property is essentially a strengthening of the requirement that the strict counter-part $\prec$ of $\preccurlyeq$ is asymmetric, by stating that for any set $\mathcal{A}' = \{C_1, \ldots, C_n\}$ of arguments, it cannot be that for all i, $C' \prec C_i$ where C' is a strict continuation of $\mathcal{A}' \backslash C_i$.

Definition 18 *(Reasonable argument orderings).* An argument ordering $\preccurlyeq$ is *reasonable* iff:

1. (i) $\forall A, B$, if A is strict and firm and B is plausible or defeasible, then $B \prec A$;
   (ii) $\forall A, B$, if B is strict and firm then $B \not\prec A$;
   (iii) $\forall A, A', B$ such that A' is a strict continuation of $\{A\}$, if $A \not\prec B$ then $A' \not\prec B$, and if $B \not\prec A$ then $B \not\prec A'$ (i.e., applying strict rules to a single argument's conclusion and possibly adding new axiom premises does not weaken, respectively strengthen, arguments).
2. Let $\{C_1, \ldots, C_n\}$ be a finite subset of $\mathcal{A}$, and for $i = 1 \ldots n$, let $C^{+\setminus i}$ be some strict continuation of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$. Then it is not the case that: $\forall i$, $C^{+\setminus i} \prec C_i$.

In Section 5.1 we give example definitions of argument orderings in terms of preorderings on the ordinary premises and defeasible rules, corresponding to the commonly used weakest and last link principles. We will then show that these argument orderings are reasonable. Henceforth, we will assume that the ordering $\preccurlyeq$ of any *(c-)SAF* is reasonable.

We now examine the implications of an argument ordering being reasonable. Under the assumption that Example 6's *SAF* is well-defined, we additionally have the arguments and attacks described in Example 7, and shown in Fig. 2(c). Recall that we have the maximal fallible sub-arguments $\{A, B'_1, B'_2\}$ of A and B, where:

- B is a strict continuation of $\{B'_1, B'_2\}$;
- A^+_1 a strict continuation of $\{B'_1, A\}$;
- A^+_2 a strict continuation of $\{B'_2, A\}$.

Assuming $\preccurlyeq$ is reasonable, then by Definition 18-2:

it cannot be that $B \prec A$, $A^+_1 \prec B'_2$ *and* $A^+_2 \prec B'_1$.

Since by assumption $B \prec A$, then it must be that either $A^+_1 \not\prec B'_2$ or $A^+_2 \not\prec B'_1$, and so A^+_1 defeats B'_2 or A^+_2 defeats B'_1. Indeed, if we refer to the preordering over defeasible rules given in Example 6, then since no rule in B'_2 is strictly stronger than a rule in A^+_1, and no rule in B'_1 is strictly stronger than a rule in A^+_2, then $A^+_1 \not\prec B'_2$, $A^+_2 \not\prec B'_1$,[7] and we obtain the defeat graph shown in Fig. 3.

In fact, the following general result can be shown:

Proposition 8. *Let A and B be arguments where B is plausible or defeasible and A and B have contradictory conclusions, and assume $\mathtt{Prem}(A) \cup \mathtt{Prem}(B)$ is c-consistent if A and B are defined as in Definition 7, that is, if they are assumed to have c-consistent premises. Then*:

1. *For all $B' \in M(B)$, there exists a strict continuation $A^+_{B'}$ of $(M(B)\setminus\{B'\}) \cup M(A)$ such that $A^+_{B'}$ rebuts or undermines B on B'.*
2. *If $B \prec A$, and $\preccurlyeq$ is reasonable, then for some $B' \in M(B)$, $A^+_{B'}$ defeats B.*

This says that if the argument ordering is reasonable, then whenever an argument B with a strict top rule rebuts (but not contrary rebuts) an argument A with a defeasible top rule but is inferior to A, we can strictly continue A into an argument defeating B.

Let us now generalise the earlier suggested counter-example to the fundamental lemma. Assume $B \in S$, S is admissible, and either: 1) B attacks A on A', $B \prec A'$, and so B does not defeat A (i.e., the example described at the beginning of this section), or; 2) A attacks B on B', $A \prec B'$, and so A does not defeat B'. The proof of Proposition 10 below then makes use of Proposition 8 to show that in neither case can A be acceptable w.r.t. S. This means that the result that *if A is acceptable w.r.t. an admissible S then $S \cup \{A\}$ is conflict free*, and hence Dung's fundamental lemma, is not under threat. Prior to Proposition 10, we state a key result for *c-SAF*s in order to show that Dung's fundamental lemma and the rationality postulates can be shown when arguments are restricted to those with consistent premises:

Proposition 9. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a c-SAF. If $A_1, \ldots, A_n$ are acceptable w.r.t. some conflict-free $E \subseteq \mathcal{A}$, then $\bigcup_{i=1}^{n} \mathtt{Prem}(A_i)$ is c-consistent.*

Proposition 10. *Let A be acceptable w.r.t. an admissible extension S of a (c-)SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$. Then $S' = S \cup \{A\}$ is conflict free.*

Proposition 10 implies that Dung's fundamental lemma holds:

[7] This is verified by the argument orderings defined on the basis of the last link principle in Section 5.1.

Proposition 11. *Let A, A' be acceptable w.r.t. an admissible extension S of a (c-)SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$. Then*:

1. $S' = S \cup \{A\}$ *is admissible.*
2. A' *is acceptable w.r.t.* S'.

We have shown that given reasonable argument orderings, a well defined *(c-)SAF* satisfies Dung's fundamental lemma. This implies that the admissible extensions of a *(c-)SAF* form a complete partial order w.r.t. set inclusion, and that every admissible extension is contained in a preferred extension. Also, given the definitions of defeat and acceptability, it is easy to see (in the same way as for Dung frameworks) that if A is acceptable w.r.t. S, then A is acceptable w.r.t. any superset of S (this result is stated as Lemma 35-1 in Appendix A). Thus, a *(c-)SAF*'s characteristic function is monotonic, implying that the grounded extension can be identified by the function's least fixed point. It is also easy to see that every stable extension is a preferred extension.

4.2. Rationality postulates for SAFs and c-SAFs under the attack definition of conflict free

As discussed in Section 1, the intermediate level of abstraction (between concrete instantiating logics and Dung's fully abstract theory) adopted by *ASPIC* [18] and *ASPIC*$^+$ [40] frameworks, allows for the formulation and evaluation of postulates [18] whose satisfaction ensure that any concrete instantiations of the frameworks fulfil some rational criteria. We now show that under the attack definition of conflict free, well-defined *SAF*s and c-*SAF*s satisfy [18]'s rationality postulates for the complete (and so by implication the grounded, preferred and stable) semantics defined in Definition 1.

Theorem 12 below states that for any argument A in a complete extension E, all sub-arguments of A are in E. Theorem 13 then states that the conclusions of arguments in a complete extension are closed under strict inference (recall that the closure $Cl_{R_s}(S)$ of S under strict rules is defined in Definition 3).

Theorem 12 *(Sub-argument closure). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a* (c-)SAF *and E an* att-*complete extension of Δ. Then for all $A \in E$: if $A' \in \mathtt{Sub}(A)$ then $A' \in E$.*

Theorem 13 *(Closure under strict rules). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a* (c-)SAF *and E an* att-*complete extension of Δ. Then $\{\mathtt{Conc}(A) | A \in E\} = Cl_{R_s}(\{\mathtt{Conc}(A) | A \in E\})$.*

Theorem 14 below, states that the conclusions of arguments in an admissible extension (and so by implication complete extension) are mutually consistent. Theorem 15 then states the mutual consistency of the strict closure of conclusions of arguments in a complete extension.

Theorem 14 *(Direct consistency). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a* (c-)SAF *and E an* att-*admissible extension of Δ. Then $\{\mathtt{Conc}(A) | A \in E\}$ is consistent.*

Theorem 15 *(Indirect consistency). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a* (c-)SAF *and E an* att-*complete extension of Δ. Then $Cl_{\mathcal{R}_s}(\{\mathtt{Conc}(A) | A \in E\})$ is consistent.*

Note that the task of showing that [18]'s consistency postulates are satisfied is simplified by the fact that a conflict free set excludes attacking arguments.

4.3. Comparison of attack and defeat definition of conflict free

Under the defeat definition of conflict free, the properties discussed in Section 4.1 are of course shown to hold for *(c-)SAF*s in the same way as for Dung frameworks. We now state the equivalence of extensions of *(c-)SAF*s under the attack and defeat definitions of conflict free.

Proposition 16. *Let Δ be a (c-)SAF. For $T \in$ {admissible, complete, grounded, preferred, stable}, E is an* att-*T extension of Δ iff E is a* def-*T extension of Δ.*

Given the previous section's results, Proposition 16 implies that [18]'s rationality postulates not only hold for *SAF*s under the defeat definition (as already shown in [40]), but also for *c-SAF*s under the defeat definition.

Corollary 17. *Let Δ be a (c-)SAF. Then Theorems 12–15 hold for the* def-*admissible and* def-*complete extensions of Δ.*

Notice that directly proving satisfaction of the consistency postulates for the defeat definition of conflict free is more involved. One must rely on Proposition 8 to show that an admissible extension contains arguments that do not defeat each

other. The trade off is that with the attack definition, proof of the fundamental lemma is more involved since one needs to first show that any argument acceptable w.r.t. an admissible extension is conflict free when included in that extension. It is the proof of *this* result that crucially depends on Proposition 8. Notice that in both cases, one needs to consider the internal structure of arguments and assume a reasonable preference ordering.

Proposition 16's equivalence begs the question as to why one should advocate the attack definition of conflict free. Firstly, a result that shows that the two different notions of conflict-freeness are (under certain assumptions) equivalent in the extensions they produce is theoretically valuable in itself. Apart from this, we have argued in Section 2 that the attack definition is conceptually more well justified. In Example 6, neither B or A defeat each other, and neither B or C defeat each other. Under the defeat definition, $\{B, A\}$ and $\{B, C\}$ are 'conflict free'. But in what meaningful sense can these sets be said to be conflict free, when they contain elements that are mutually inconsistent? Consider then [7]'s example that purports to illustrate violation of the consistency postulates by approaches augmenting Dung frameworks with preferences. An expert argues (A) that a given violin is a Stradivarius and therefore expensive. A three-year old child's argument B then states that it is not a Stradivarius. According to [7], B attacks A but A does not attack B, and A is preferred over B since the expert is more reliable than the child. [7] observe that the unique *PAF*-extension $\{A, B\}$ violates the consistency postulate. In Section 6.2 we demonstrate that the problem does not arise if all arguments that can be constructed are taken into account (the expert can use a sub-argument A' of A that defeats B so that $\{A, B\}$ is not admissible), illustrating that the problem has more to do with imperfect reasoners. However, we also note that the attack definition of conflict free is more tolerant of imperfect reasoning. Without taking into account all constructible arguments, $\{A, B\}$ is of course *not* conflict free and so not a *PAF*-extension, and so consistency is not violated.

To conclude, given our advocacy of the attack definition of conflict free, we henceforth assume any extension of a *(c-)SAF* to be attack conflict free, and thus will henceforth (for a given semantics T) refer to a 'T extension' rather than an '*att-T* extension'. However, for the results shown in the next section, we will indicate, when appropriate, that the results also hold under the defeat definition of conflict free.

5. Instantiating structured argumentation frameworks

We have modified *ASPIC*$^+$ in two ways: we have additionally defined *c-SAF*s whose arguments must be built on mutually consistent premises, and motivated an alternative attack definition of conflict free sets. We have shown that properties and postulates hold for well defined *SAF*s and *c-SAF*s with argument preference orderings that are *reasonable*. In this section we study various ways to instantiate the *ASPIC*$^+$ framework. Section 5.1 consider ways of 'instantiating' preference orderings over arguments in terms of preorderings over defeasible rules and ordinary premises. We show that the defined argument orderings are reasonable. In Section 5.2 we extend with preferences Amgoud & Besnard's approach to structured argumentation [2,3] based on Tarski's notion of an abstract logic. We then combine the extended abstract logic approach with *ASPIC*$^+$. In Section 5.3 we define classical logic instantiations of *c-SAF*s, and show an equivalence between one such instantiation and Brewka's Preferred Subtheories [16].

5.1. Weakest and last link preference relations

In [40], a strict argument ordering $\prec$ is defined on the basis of two preorderings $\leqslant$ on $\mathcal{R}_d$ and $\leqslant'$ on $\mathcal{K}_p$ under the well known *weakest*-link [16,21] and *last*-link [29,44] principles.[8] Intuitively, $B \prec A$ is defined by separate set comparisons of the defeasible rules in B and A, and the ordinary premises in B and A. Then $B \prec A$ by the weakest link principle if:

1. from amongst *all the* defeasible rules in B there exists a rule which is weaker than (strictly less than according to $\leqslant$) *all the* defeasible rules in A, *and*
2. from amongst all the ordinary premises in B there is an ordinary premise which is weaker (strictly less than according to $\leqslant'$) all the ordinary premises in A.

Then $B \prec A$ by the last link principle if the above set comparison (henceforth referred to as the `Elitist` comparison) on defeasible rules is now applied only to the last defeasible rules in B and A (recall the definition of `LastDefRules` in Definition 5); i.e., 'all the last' replaces 'all the' in 1). If there are no defeasible rules in B and A, then only the ordinary premises are compared, and so $B \prec A$ by the last link principle if 2) holds.

In this paper we address two limitations of the way argument orderings are defined in [40]. Firstly, in [40], strict preferences over arguments ($\prec$) are defined directly, in terms of the above set comparisons. However, to comply with the definition of the general *ASPIC*$^+$framework, we now define the non-strict ordering $\preccurlyeq$, letting the strict counterpart $\prec$ be then defined in the usual way. Unlike [40], we can then identify equivalence classes of arguments under $\preccurlyeq$ (recall that $A \approx B$ iff $A \preccurlyeq B$ and $B \preccurlyeq A$). To directly define $\preccurlyeq$ under the weakest or last link principles, in turn requires that we define non-strict set comparisons over the defeasible rules/axiom premises (in which 'less than or equal' replaces 'strictly less than'

[8] Note that the cited papers make use of the principles without explicitly naming them as such.

in 1) and 2) above). The second limitation we address is an anomaly in [40]'s definition of the weakest link principle, and is discussed at further length after Definition 21 below.

Finally, we broaden the range of instantiations of $\preccurlyeq$ considered in [40], by allowing for an alternative comparison of the (last) defeasible rules and ordinary premises in B and A. Specifically, we provide an alternative interpretation of the weakest and last link principles based on an alternative set comparison (sometimes referred to as the `Democratic` comparison [21]). In what follows, we provide a general definition of a set comparison $\trianglelefteq_{\mathtt{s}}$ that can be applied to defeasible rules and premises, and which is then parameterised according to the `Elitist` and `Democratic` comparisons (i.e., $\mathtt{s} = \mathtt{Eli}$ and `Dem` respectively):

Definition 19 *(Orderings $\trianglelefteq_{\mathtt{s}}$).* Let Γ and Γ' be finite sets.[9] Then $\trianglelefteq_{\mathtt{s}}$ is defined as follows:

1. If $\Gamma = \emptyset$ then $\Gamma \not\trianglelefteq_{\mathtt{s}} \Gamma'$;
2. If $\Gamma' = \emptyset$ and $\Gamma \neq \emptyset$ then $\Gamma \trianglelefteq_{\mathtt{s}} \Gamma'$;
   else:
3. assuming a preordering $\leqslant$ over the elements in $\Gamma \cup \Gamma'$, then if $\mathtt{s} = \mathtt{Eli}$:
   $\Gamma \trianglelefteq_{\mathtt{Eli}} \Gamma'$ if $\exists X \in \Gamma$ s.t. $\forall Y \in \Gamma'$, $X \leqslant Y$.
   else:
4. assuming a preordering $\leqslant$ over the elements in $\Gamma \cup \Gamma'$, then if $\mathtt{s} = \mathtt{Dem}$:
   $\Gamma \trianglelefteq_{\mathtt{Dem}} \Gamma'$ if $\forall X \in \Gamma$, $\exists Y \in \Gamma'$, $X \leqslant Y$.

The strict counterpart of $\trianglelefteq_{\mathtt{s}}$ is defined in the usual way: $\Gamma \triangleleft_{\mathtt{s}} \Gamma'$ iff $\Gamma \trianglelefteq_{\mathtt{s}} \Gamma'$ and $\Gamma' \not\trianglelefteq_{\mathtt{s}} \Gamma$.

Note that for any sets of defeasible rules/ordinary premises S and S', we intuitively want that:

1) if S is the empty set, it cannot be that $S \triangleleft S'$;
2) if S' is the empty set, it must be that $S \triangleleft S'$ for any non-empty S.

Hence the above definition explicitly imposes that any set comparison $\trianglelefteq_{\mathtt{s}}$ satisfies these properties since one cannot assume them to be satisfied for every possible `s`. For example, the `Democratic` comparison does not in general satisfy these properties.[10]

Definition 20 *(Last-link principle).* Let $\mathtt{s} \in \{\mathtt{Eli}, \mathtt{Dem}\}$. Then $B \preccurlyeq A$ under the last-link principle iff

1. $\mathtt{LastDefRules}(B) \trianglelefteq_{\mathtt{s}} \mathtt{LastDefRules}(A)$; or
2. $\mathtt{LastDefRules}(B) = \emptyset$, $\mathtt{LastDefRules}(A) = \emptyset$, and $\mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq_{\mathtt{s}} \mathtt{Prem}_{\mathtt{p}}(A)$.

Definition 21 *(Weakest-link principle).* Let $\mathtt{s} \in \{\mathtt{Eli}, \mathtt{Dem}\}$. Then $B \preccurlyeq A$ under the weakest-link principle iff:

1. If both B and A are strict, then $\mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq_{\mathtt{s}} \mathtt{Prem}_{\mathtt{p}}(A)$, else;
2. If both B and A are firm, then $\mathtt{DefRules}(B) \trianglelefteq_{\mathtt{s}} \mathtt{DefRules}(A)$, else;
3. $\mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq_{\mathtt{s}} \mathtt{Prem}_{\mathtt{p}}(A)$ and $\mathtt{DefRules}(B) \trianglelefteq_{\mathtt{s}} \mathtt{DefRules}(A)$.

Notice that [40]'s definition of the weakest link principle implies an anomaly that is corrected here. [40]'s definition (which, recall, defines $\prec$ directly in terms of a directly defined ordering $\triangleleft$) implies that if both B and A are strict (contain no defeasible rules), then $B \prec A$ if there are ordinary premises in B that are $\triangleleft$ the ordinary premises in A. However, if both B and A are firm (contain no ordinary premises), then it is not the case that $B \prec A$ if there are defeasible rules in B that are $\triangleleft$ the defeasible rules in A. Thus there is an asymmetry in the way that premises and defeasible rules are compared. In the above definition, the weakest link is defined so that the defeasible rules and ordinary premises are treated in the same way.

Example 18 *(Example 4 continued).* Given:

- $r \Rightarrow q \leqslant a \Rightarrow p$;
- $\neg r <' r$; $\neg a \approx' r$; $\sim s <' \neg r$

[9] Notice that it suffices to restrict $\triangleleft$ to finite sets since *ASPIC*$^+$ arguments are defined as finite (in Definition 5) and so their ordinary premises/defeasible rules must be finite.

[10] Since if $S = \emptyset$, $S' \neq \emptyset$, then trivially: $\forall X \in S$, $\exists Y \in S'$ *s.t.* $X \leqslant Y$ and it is not the case that $\forall Y \in S'$, $\exists X \in S$ *s.t.* $Y \leqslant X$.

on defeasible rules and ordinary premises in Example 1, and employing the abbreviations DR for DefRules and LDR for LastDefRules, we have:

$$\mathtt{DR}(A) = \mathtt{LDR}(A) = \{a \Rightarrow p\}, \qquad \mathtt{Prem}_{\mathtt{p}}(A) = \{a\};$$
$$\mathtt{DR}(A') = \mathtt{LDR}(A') = \emptyset, \qquad \mathtt{Prem}_{\mathtt{p}}(A) = \{a\};$$
$$\mathtt{DR}(B) = \mathtt{LDR}(B) = \{\sim s \Rightarrow t, r \Rightarrow q\}, \qquad \mathtt{Prem}_{\mathtt{p}}(B) = \{\sim s, r\};$$
$$\mathtt{DR}(B_2) = \mathtt{LDR}(B_2) = \emptyset, \qquad \mathtt{Prem}_{\mathtt{p}}(B_2) = \{r\};$$
$$\mathtt{DR}(C) = \mathtt{LDR}(C) = \emptyset, \qquad \mathtt{Prem}_{\mathtt{p}}(C) = \{\neg r\}.$$

Then:

- $\mathtt{LDR}(B) \trianglelefteq_{\mathtt{Eli}} \mathtt{LDR}(A)$, $\mathtt{LDR}(A) \not\trianglelefteq_{\mathtt{Eli}} \mathtt{LDR}(B)$ and so $B \preccurlyeq A$, $A \not\preccurlyeq B$, hence $B \prec A$ under the last-link principle. $\mathtt{DR}(B) \trianglelefteq_{\mathtt{Eli}} \mathtt{DR}(A)$, but $\mathtt{Prem}_{\mathtt{p}}(B) \not\trianglelefteq_{\mathtt{Eli}} \mathtt{Prem}_{\mathtt{p}}(A)$, so $B \not\preccurlyeq A$ hence $B \not\approx A$ under the weakest-link principle.
- $\mathtt{LDR}(B) \not\trianglelefteq_{\mathtt{Dem}} \mathtt{LDR}(A)$ and $\mathtt{Prem}_{\mathtt{p}}(B) \not\trianglelefteq_{\mathtt{Dem}} \mathtt{Prem}_{\mathtt{p}}(A)$, and so $B \not\preccurlyeq A$, hence $B \not\approx A$ under the last or weakest-link principle.
- $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq_{\mathtt{Eli}} \mathtt{Prem}_{\mathtt{p}}(B_2)$, $\mathtt{Prem}_{\mathtt{p}}(B_2) \not\trianglelefteq_{\mathtt{Eli}} \mathtt{Prem}_{\mathtt{p}}(C)$ and so $C \prec B_2$ under the last or weakest-link principle.
- $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq_{\mathtt{Dem}} \mathtt{Prem}_{\mathtt{p}}(B_2)$, $\mathtt{Prem}_{\mathtt{p}}(B_2) \not\trianglelefteq_{\mathtt{Dem}} \mathtt{Prem}_{\mathtt{p}}(C)$ so $C \prec B_2$ under the last or weakest-link principle.
- $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq_{\mathtt{Dem}} \mathtt{Prem}_{\mathtt{p}}(A')$, $\mathtt{Prem}_{\mathtt{p}}(A') \trianglelefteq_{\mathtt{Dem}} \mathtt{Prem}_{\mathtt{p}}(C)$, and so $C \preccurlyeq A'$, $A' \preccurlyeq C$, hence $C \approx A'$ under the last or weakest-link principle.

A natural question to ask is whether comparisons other than Democratic or Elitist can be employed when defining $\trianglelefteq_{\mathtt{s}}$ in Definition 19, so allowing reference to a larger range of comparisons s in the definitions of the last and weakest link principles. In what follows we identify a class of such comparisons, by specifying properties that the defined set ordering $\trianglelefteq_{\mathtt{s}}$ satisfies.

Definition 22 *(Inducing reasonable orderings).* $\trianglelefteq_{\mathtt{s}}$ is said to *reasonable inducing* if:

1. $\trianglelefteq_{\mathtt{s}}$ is transitive;
2. for any $\mathtt{kr} \in \{\mathtt{LastDefRules}, \mathtt{DefRules}, \mathtt{Prem}_{\mathtt{p}}\}$, for all arguments $B_1, \ldots, B_n, A$ such that $\bigcup_{i=1}^{n} \mathtt{kr}(B_i) \triangleleft \mathtt{kr}(A)$, it holds that:
   (a) for some $i = 1 \ldots n$, $\mathtt{kr}(B_i) \trianglelefteq \mathtt{kr}(A)$; and
   (b) for some $i = 1 \ldots n$, $\mathtt{kr}(A) \not\trianglelefteq \mathtt{kr}(B_i)$.

We now show that the last and weakest link orderings $\preccurlyeq$ are reasonable under the assumption of any $\trianglelefteq_{\mathtt{s}}$ that is reasonable inducing.

Proposition 19. *Let $\preccurlyeq$ be defined according to the last-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is reasonable inducing. Then $\preccurlyeq$ is reasonable.*

Proposition 20. *Let $\preccurlyeq$ be defined according to the weakest-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is reasonable inducing. Then $\preccurlyeq$ is reasonable.*

The following propositions therefore imply that the last and weakest link orderings $\preccurlyeq$ defined in Definitions 20 and 21 are reasonable.

Proposition 21. *$\trianglelefteq_{\mathtt{Eli}}$ is reasonable inducing.*

Proposition 22. *$\trianglelefteq_{\mathtt{Dem}}$ is reasonable inducing.*

Finally, note that if $\trianglelefteq_{\mathtt{s}}$ is transitive, then the strict counterparts $\prec$ of the weakest and last link orderings $\preccurlyeq$ are strict partial orders:

Proposition 23. *Let $\preccurlyeq$ be defined according to the last-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is transitive. Then $\prec$ is a strict partial order.*

Proposition 24. *Let $\preccurlyeq$ be defined according to the weakest-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is transitive. Then $\prec$ is a strict partial order.*

5.2. Reconstructing and extending the abstract logic approach as an instance of ASPIC$^+$

In [2,3], Amgoud & Besnard present an abstract approach to defining the structure of arguments and attacks, based on Tarski's notion of an *abstract logic*.

Definition 23 *(Abstract logic).* An abstract logic is a pair $(\mathcal{L}, Cn)$, where $\mathcal{L}$ is a language and the consequence operator Cn is a function from $2^{\mathcal{L}}$ to $2^{\mathcal{L}}$ satisfying the following conditions for all $X \subseteq \mathcal{L}$:

1. $X \subseteq Cn(X)$.
2. $Cn(Cn(X)) = Cn(X)$.
3. $Cn(X) = \bigcup_{Y \subseteq_f X} Cn(Y)$.
4. $Cn(\{p\}) = \mathcal{L}$ for some $p \in \mathcal{L}$.
5. $Cn(\emptyset) \neq \mathcal{L}$.

Here $Y \subseteq_f X$ means that Y is a finite subset of X. A set $X \subseteq \mathcal{L}$ is defined as *consistent* if $Cn(X) \neq \mathcal{L}$, and as inconsistent otherwise.

Amgoud & Besnard [2] note that the following properties hold:

6. If $X \subseteq X'$ then $Cn(X) \subseteq Cn(X')$ (*monotonicity*).
7. If $Cn(X) = Cn(X')$ then $Cn(X \cup Y) = Cn(X' \cup Y)$.

[2] also restricts its focus to so-called *adjunctive* abstract logics:

8. $\forall x, y \in \mathcal{L}$ such that $Cn(\{x, y\}) \neq Cn(\{x\}), Cn(\{x, y\}) \neq Cn(\{y\})$, $\exists z$ such that $z \neq x$, $z \neq y$ and $Cn(\{z\}) = Cn(\{x, y\})$.[11]

They then define arguments and various kinds of attack relations, and investigate consistency properties of various types of attack relations when instantiating Dung's framework with arguments and attacks. We discuss this part of their work in Section 6. We repeat here [3]'s notion of an undermining attack.[12] We also extend their approach to accommodate a preordering over the formulae in an abstract logic theory.

Definition 24 *(Arguments and attacks in abstract logics).* Let $(\mathcal{L}, Cn)$ be an abstract logic and $(\Sigma, \leqslant)$ a theory in $(\mathcal{L}, Cn)$, where $\Sigma \subseteq \mathcal{L}$ and $\leqslant$ a preordering over Σ:

- an *AL-argument* is a pair (X, p) such that: 1) $X \subseteq \Sigma$; 2) X is consistent; 3) $p \in Cn(X)$; 4) no proper subset of X satisfies (1–3).
- (X, p) *AL-undermines* (Y, q) if there exists a $q' \in Y$ such that $\{p, q'\}$ is inconsistent.

We formally define the notion of an *ASPIC*$^+$ argumentation theory based on an abstract logic with preferences. This involves defining the set of strict rules in terms of the abstract-logic's consequence notion but also relating the $^-$ relation to the Tarskian notion of consistency. Note that the latter does not allow for defining asymmetric contrary relations, and so we have to assume that *ASPIC*$^+$'s $^-$ relation is symmetric. Next, two conditions are needed to relate the $^-$ relation to the Tarskian notion of consistency. Firstly, if two formulae are contradictories of each other then they are jointly inconsistent. Secondly, if two formulae are jointly inconsistent, then each of them has a consequence that is a contradictory of the other. Also, a knowledge base will consist of the elements of an abstract logic theory as ordinary premises, while the argument ordering will be defined in terms of a preordering on the abstract logic theory. Finally, to avoid confusion we henceforth refer to the abstract logic notion of consistency as 'AL-consistency'.

Definition 25 *(AT and c-SAF based on abstract logic with preferences).* Let $(\mathcal{L}', Cn)$ be an abstract logic and $(\Sigma, \leqslant')$ a theory in $(\mathcal{L}', Cn)$. An *abstract logic (AL) argumentation theory* based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$, is a pair $(AS, \mathcal{K})$ such that AS is an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n)$ based on $(\mathcal{L}', Cn)$, where:

1. $\mathcal{L} = \mathcal{L}'$;
2. $\mathcal{R}_d = \emptyset$, and for all finite $S \subseteq \mathcal{L}$ and $p \in \mathcal{L}$, $S \rightarrow p \in \mathcal{R}_s$ iff $p \in Cn(S)$;
3. $^-$ is defined such that:
   (a) if $\varphi \in \overline{\psi}$ then $\psi \in \overline{\varphi}$;
   (b) if $\varphi \in \overline{\psi}$ then $\{\varphi, \psi\}$ is AL-inconsistent;

[11] For example, classical logic is adjunctive because of the conjunction connective.
[12] [3] call undermining "undercutting" but to be consistent with *ASPIC*$^+$'s terminology we rename it to 'undermining'.

(c) if $\{\varphi, \psi\}$ is AL-inconsistent then there exists a $\varphi' \in Cn(\{\varphi\})$ such that $\varphi' \in \overline{\psi}$;
(d) $\overline{\varphi}$ is non-empty for all φ.

$\mathcal{K}$ is a knowledge base such that $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \Sigma$.

$(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ is the *c-SAF* based on $(AS, \mathcal{K})$, as defined in Definition 11 and where $\preccurlyeq$ is defined in terms of $\leqslant'$ as in Section 5.1. We also say that $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ is the *c-SAF* based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$.

We can then show that a *c-SAF* based on an abstract logic with preferences is well defined:

Proposition 25. *A c-SAF based on an AL argumentation theory is closed under contraposition, axiom consistent, c-classical, and well-formed.*

Since $\preccurlyeq$ is reasonable, Proposition 25 implies that all the results and rationality postulates in Sections 4.1 and 4.2 hold for *c-SAF*s based on an abstract logic with preferences. However, note that these *c-SAF*s are instantiated by *ASPIC*$^+$ arguments and undermining attacks (since rebuts and undercuts only apply to defeasible rules). The question naturally arises as to whether they are equivalent to the AL arguments and undermining attacks in Definition 24. We first show that the attacks are indeed equivalent. To do so, we define the notion of an *AL-c-SAF*:

Definition 26. Let an ASPIC$^+$*-AL-undermining* attack be defined in the same way as an *ASPIC*$^+$undermining attack, with '$\texttt{Conc}(A) \vdash -\varphi$' replacing '$\texttt{Conc}(A) \in \overline{\varphi}$' in Definition 8.

Then an *AL-c-SAF* defined by $(AS, \mathcal{K})$ is defined as in Definition 25, with '$(X, Y) \in \mathcal{C}$ iff X *ASPIC*$^+$-AL-undermines Y' replacing '$(X, Y) \in \mathcal{C}$ iff X attacks Y' in Definition 11.

Given Lemma 42 in Section A.5, which shows that $\texttt{Conc}(A) \vdash -\varphi$ iff $\{\texttt{Conc}(A), \varphi\}$ is AL-inconsistent, then the following result shows that *ASPIC*$^+$'s undermining attacks faithfully reconstruct abstract logic undermining attacks:

Proposition 26. [13] *Let $(AS, \mathcal{K})$ be based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Let Δ_1 be the c-SAF defined by $(AS, \mathcal{K})$ and $\leqslant'$, and Δ_2 the AL-c-SAF defined by $(AS, \mathcal{K})$ and $\leqslant'$. Then, for $T \in$ {complete, grounded, preferred, stable}, E is a T extension of Δ_1 iff E is a T extension of Δ_2.*

Now, observe that we do *not* have an equivalence between *ASPIC*$^+$ arguments and AL arguments, because the latter imposes a subset minimality condition on the premises. This condition is not imposed on *ASPIC*$^+$ arguments in Definition 5, and neither is it implied by the definition of $\mathcal{R}_s$ in Definition 25. Consider the following counter-example. Given $q \in Cn(\{p\})$, $s \in Cn(\{p, r\})$ and $q \in Cn(\{s\})$, we obtain $\mathcal{R}_s = \{p \rightarrow q; p, r \rightarrow s; s \rightarrow q\}$. Then we have the strict arguments $\{p\} \vdash q$ and $\{p, r\} \vdash q$ where the latter is not subset minimal.

In general, minimality of premise sets is undesirable. Suppose a defeasible rule $p \Rightarrow q$ and a strict rule $p, r \rightarrow q$: then we clearly do not want to rule out an argument for q with premises p and r, since it could well be stronger than the defeasible argument. However, since the *ASPIC*$^+$ arguments defined by an AL argumentation theory are strict, we define here the notion of premise minimal *ASPIC*$^+$ arguments and show an equivalence with AL arguments.

Definition 27 *(Premise minimal ASPIC*$^+$ *arguments).* Let for any argument A, A_- be any argument such that $\texttt{Prem}(A_-) \subseteq \texttt{Prem}(A)$ and $\texttt{Conc}(A_-) = \texttt{Conc}(A)$. Given a set of *ASPIC*$^+$ arguments $\mathcal{A}$, let $\mathcal{A}^- = \{A \in \mathcal{A} \mid$ there is no $A_- \in \mathcal{A}$ such that $\texttt{Prem}(A_-) \subset \texttt{Prem}(A)\}$ be the premise minimal arguments in $\mathcal{A}$.

Proposition 27. *Let $(AS, \mathcal{K})$ be based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Then A is a c-consistent premise minimal argument on the basis of $(AS, \mathcal{K})$ iff $(\texttt{Prem}(A), \texttt{Conc}(A))$ is an abstract logic argument on the basis of $(\Sigma, \leqslant')$.*

We can then show that for *c-SAF*s *and SAF*s, when restricting consideration to arguments with minimal premise sets, the conclusions of arguments in complete extensions remains unchanged, under the assumption that an argument cannot be strengthened by just adding premises. The latter is formulated by requiring that if B is not strictly preferred to A then B is not strictly preferred to A_- (since if B were strictly preferred to A_- this would imply that A_- has been strengthened by adding premises to obtain A).

Proposition 28. *Let Δ be the (c)-SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ defined on the basis of an AT for which $\preccurlyeq$ is defined such that for any $A \in \mathcal{A}$, if $A \not\preccurlyeq B$ then $A_- \not\preccurlyeq B$.*

[13] Note that in Appendix A, the proof of Proposition 26 shows that the result holds under both attack *and* defeat definitions of conflict free.

Let Δ^- be the premise minimal (c)-SAF $(\mathcal{A}^-, \mathcal{C}^-, \preccurlyeq^-)$ where:

- $\mathcal{A}^-$ *is the set of premise minimal arguments in* $\mathcal{A}$.
- $\mathcal{C}^- = \{(X,Y) | (X,Y) \in \mathcal{C}, X, Y \in \mathcal{A}^-\}$.
- $\preccurlyeq^- = \{(X,Y) | (X,Y) \in \preccurlyeq, X, Y \in \mathcal{A}^-\}$.

Then for $T \in$ {complete, grounded, preferred, stable}, E is a T extension of Δ iff E' is a T extension of Δ^-, where $E' \subseteq E$ and $\bigcup_{X \in E} \mathtt{Conc}(X) = \bigcup_{Y \in E'} \mathtt{Conc}(Y)$.

Note that although the above proposition assumes the attack definition of conflict free, it immediately follows from Proposition 16 that Proposition 28 also holds if the defeat definition of conflict free is assumed.

Corollary 29. *Given Δ and Δ^- as defined in Proposition 28*:

1. *φ is a T credulously (sceptically) justified conclusion of Δ iff φ is a T credulously (sceptically) justified conclusion of Δ^-.*
2. *Δ^- satisfies the postulates* closure under strict rules, direct consistency, indirect consistency *and* sub-argument closure.

The assumption that arguments are not strengthened by adding premises is not satisfied by all ways of defining $\preccurlyeq$. Consider a *c-SAF* $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ defined by an AL argumentation theory, where $\preccurlyeq$ is defined on the basis of the democratic comparison $\trianglelefteq_{\mathtt{Dem}}$, and suppose arguments A_-, A and B such that $\mathtt{Prem}(A_-) = \{p\}$, $\mathtt{Prem}(A) = \{p, q\}$, $\mathtt{Prem}(B) = \{r\}$, and assume the preordering on the premises is $p \leqslant' r$. Then $\{p, q\} \not\trianglelefteq_{\mathtt{Dem1}} \{r\}$, but $\{p\} \trianglelefteq_{\mathtt{Dem1}} \{r\}$, and so it is easy to verify that $A \not\prec B$, but $A_- \prec B$. However, the assumption is satisfied by the elitist $\trianglelefteq_{\mathtt{Eli}}$:

Proposition 30. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be defined by an AL argumentation theory, where $\preccurlyeq$ is defined under the weakest or last link principles, based on the set comparison $\trianglelefteq_{\mathtt{Eli}}$. Then $\forall A, B \in \mathcal{A}$, $\forall A_- \in \mathcal{A}$, if $A \not\prec B$ then $A_- \not\prec B$.*

In conclusion:

Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be the *c-SAF* based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$,
as defined in Definition 25.

We have shown that all results and rationality postulates hold for Δ. We have also shown that the AL undermining attacks and $\mathcal{C}$ are equivalent in the complete extensions that they generate, and the AL arguments and premise minimal *ASPIC*$^+$ arguments in $\mathcal{A}$ are equivalent.

Furthermore, Proposition 28 and Corollary 29 then imply that:

Remark 31. 1) We have combined [2,3]'s abstract logic approach to argumentation (which assumes no preference relation over Σ) with the *ASPIC*$^+$ framework, in that the justified conclusions of a Dung framework instantiated by AL arguments and AL undermining attacks, are exactly those of Δ. We have also shown that Section 4.2's rationality postulates hold for Amgoud & Besnard's approach.

2) Given that we have extended the abstract logic approach to accommodate preferences, consider a preference-based argumentation framework (see Section 2.1) $\Gamma = (\mathcal{A}', \mathcal{C}', \preccurlyeq')$ defined by $(\Sigma, \leqslant')$ and $(\mathcal{L}', Cn)$, where $\mathcal{A}'$ and $\mathcal{C}'$ are the AL arguments and AL undermining attacks. Then, under the assumption that $\preccurlyeq'$ does not strengthen arguments when adding premises, the justified conclusions of Γ (specifically the Dung framework instantiated by arguments and defeats defined by Γ) are exactly those of Δ. This assumption is satisfied when defining $\preccurlyeq'$ under the last or weakest link principles, based on Definition 19's elitist set comparison that utilises the preordering $\leqslant'$ over formulae in Σ. We have also shown that Section 4.2's rationality postulates hold for Amgoud & Besnard's approach extended with preferences.

We conclude by observing that Amgoud & Besnard investigate the consistency of extensions of a Dung framework instantiated by the arguments and attacks defined by an abstract logic. Specifically, they consider whether for any extension the union of the premises of the extension's arguments is AL consistent. We now show that for a *c-SAF* based on an abstract logic this is equivalent to indirect consistency, and then refer to this result in Section 6 in which we compare *ASPIC*$^+$ and the abstract logic approach.

Proposition 32. *Let Δ be the c-SAF based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Then for any complete extension E of Δ: $S = \{\phi | \phi \in \mathtt{Prem}(A), A \in E\}$ is AL-inconsistent iff $S' = Cl_{\mathcal{R}_s}(\{\mathtt{Conc}(A) | A \in E\})$ is inconsistent.*

5.3. Classical logic instances of the ASPIC$^+$ framework

The previous section's results allow us to reconstruct classical logic approaches to argumentation as a special case of *ASPIC*$^+$, and in so doing extend these approaches with preferences. We also prove a relation with Brewka's preferred subtheories.

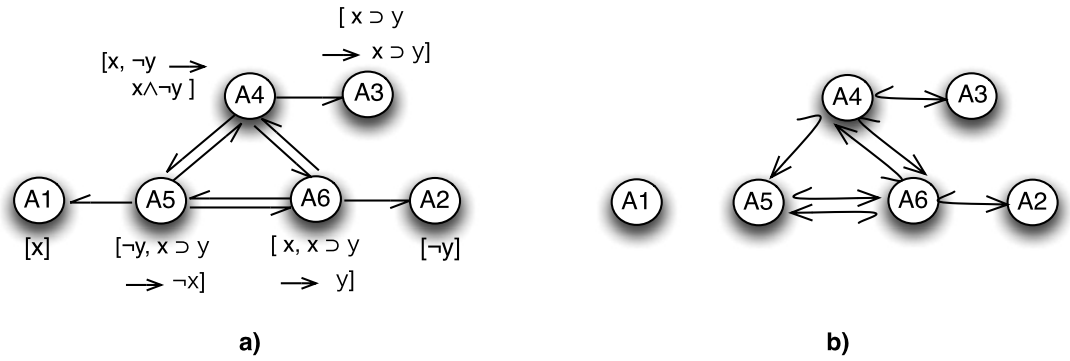

Fig. 4. Classical Logic argumentation: attack graph (a) and defeat attack (b).

5.3.1. Defining classical logic instantiations of ASPIC$^+$

Much recent work on structured argumentation formalises arguments as minimal classical consequences from consistent and finite premise sets in standard propositional or first-order logic [5,12,13,26]. Since classical logic can be specified as a Tarskian abstract logic $(\mathcal{L}', Cn)$, where $\mathcal{L}'$ is a standard propositional or first-order language and Cn the classical consequence relation, a *classical argumentation theory* and its *c-SAF* based on $(\mathcal{L}', Cn)$ and an ordered theory $(\Sigma, \leqslant')$, is defined as in Definition 25. The ordering $\leqslant'$ on Σ and thus the ordinary premises, allows us to reconstruct classical logic approaches that additionally consider preferences (e.g., [5]). It is then easy to verify that if $^-$ is defined as classical negation, then all four conditions in Definition 25-(3) are satisfied.

Amongst the above-cited works on classical argumentation, [5] and [26] adopt a Dung-style semantics, where only [5] considers preferences. Let us first consider [26]. They define seven alternative notions of attack and investigate their properties, including the rationality postulates of [18] studied in this paper. They show that the only two attack relations that are 'well behaved', in the sense that they satisfy consistency postulate for all the semantics, are the so-called 'direct undercuts' and 'direct defeaters':

- *Y directly undercuts X* if $\texttt{Conc}(Y) \equiv \neg p$ for some $p \in \texttt{Prem}(X)$.
- *Y directly defeats X* if $\texttt{Conc}(Y) \vdash_c \neg p$ for some $p \in \texttt{Prem}(X)$.

Although our undermining attacks are not among [26]'s seven notions of attack, it can be shown that our undermining attacks are equivalent to their direct undercuts and defeats in that the complete extensions generated are the same. For direct defeats, this result is shown by Proposition 26. For direct undercuts, it suffices to adapt the proof of Proposition 26, showing that: (1) if Y undermines X, then letting $\texttt{Conc}(Y) = q$, by the symmetry of classical negation $q = \neg p$ for some $p \in \texttt{Prem}(X)$ and so $\texttt{Conc}(Y)) \equiv \neg p$; (2) if Y directly undercuts X then Y directly defeats X, and so as already shown, Y undermines X. These equivalences and [26]'s negative results for their remaining five notions of attack justify why *ASPIC*$^+$ does not model these five notions.

It follows from the above, and the results and discussion in Section 5.2, that we have reconstructed and extended with preferences, [26]'s variants with direct undercut and direct defeat, and shown that [18]'s postulates are satisfied for classical logic approaches with preferences (recall that [26]'s other variants violate the consistency postulate even without preferences).

Example 33. Let the ordinary premises be the set $\Sigma = \{x, \neg y, x \supset y\}$ ($\supset$ denotes material implication) and assume $x >' \neg y, x >' x \supset y$. The attack graphs is shown in Fig. 4(a). Under either the weakest or last link principles, and assuming either the elitist or democratic comparisons, $A_5 \prec A_1$ and so A_5 does not defeat A_1. Note also that A_5 attacks A_4 on A_1, and so A_5 does not defeat A_4. The defeat graph is shown in Fig. 4(b).

We obtain $E'_1 = \{A_1, A_4, A_2\}$ and $E_2 = \{A_1, A_6, A_3\}$, where by satisfaction of the closure under strict rules postulate, $E_1 = E'_1$ extended with arguments concluding classical consequences of $\{x, \neg y, x\neg y\}$ is a preferred/stable extension, and $E_2 = E'_2$ extended with arguments concluding classical consequences of $\{x, y, x \supset y\}$ is a preferred/stable extension.

The above example shows how preferences arbitrate in favour of the sceptically justified conclusion x over $\neg x$. Indeed, we argue that extending classical logic approaches with preferences is of particular importance, given that (as shown in [20,26]) the preferred/stable extensions generated from a Dung framework instantiated by arguments and direct undercuts or defeats, simply correspond to the maximal consistent subsets of the theory Σ from which the arguments are defined.[14] Intuitively, one would expect this correspondence given that classical logic does not provide any logical machinery for

[14] In the sense that the union of formulae in the supports of arguments in each preferred/stable extension is a maximal consistent subset of Σ.

arbitrating conflicts (in contrast with the use of undercuts and negation as failure in non-monotonic logics (as discussed in Sections 2.3 and 3.2). One must therefore resort to some meta-logical mechanism, such as preferences, if argumentation is to be usefully deployed in resolving inconsistencies in a classical-logic setting.

We conclude by noting that [5] make use of preferences to determine the success of two of [26]'s variants of attack, and show that this leads to violation of the consistency postulates. We will discuss this in detail in Section 6.

5.3.2. Brewka's preferred subtheories as an instance of the ASPIC$^+$ *framework*

Brewka's *preferred subtheories* [16] models the use of an ordering over a classical propositional or first-order theory Γ, in order to resolve inconsistencies. It has therefore been used to both formalise default reasoning and belief revision [17].

Definition 28. A default theory Γ is a tuple $(\Gamma_1, \ldots, \Gamma_n)$, where each Γ_i is a set of formulae in a classical first-order language $\mathcal{L}'$. A preferred subtheory is a set $\Sigma = \Sigma_1 \cup \cdots \cup \Sigma_n$ such that for $i = 1 \ldots n$, $\Sigma_1 \cup \cdots \cup \Sigma_i$ is a maximal (under set inclusion) consistent subset of $\Gamma_1, \ldots, \Gamma_i$.

Intuitively, a preferred subtheory is obtained by taking a maximal under set inclusion consistent subset of Γ_1, extending this with a maximal consistent subset of Γ_2, extending this with a maximal consistent subset of Γ_3, and so on. We can reconstruct preferred subtheories as an instance of the *ASPIC*$^+$ framework.

Definition 29. Let Γ be a default theory $(\Gamma_1, \ldots, \Gamma_n)$, and $\forall \alpha, \beta \in \Gamma$, $(\alpha, \beta) \in \preccurlyeq'$ iff $\alpha \in \Gamma_i, \beta \in \Gamma_j$, $i \geqslant j$. Let Δ be the *c-SAF* $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ based on $(\mathcal{L}', Cn)$ and $(\Gamma, \preccurlyeq')$ as described in Section 5.3.1 (with Γ replacing Σ), and where $\preccurlyeq$ is defined under the weakest or last link principle, and on the basis of the $\trianglelefteq_{\texttt{Eli}}$ set comparison. We say that Δ is the *c-SAF* corresponding to Γ.

Theorem 34. *Let* $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ *be a c-*SAF *corresponding to a default theory* Γ*, and for any* $\Sigma \subseteq \Gamma$*, let* $\texttt{Args}(\Sigma) \subseteq \mathcal{A}$ *be the set of all arguments with premises taken from* Σ*. Then*:

1) *If* Σ *is a preferred subtheory of* Γ*, then* $\texttt{Args}(\Sigma)$ *is a stable extension of* $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$*.*
2) *If E is a stable extension of* $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$*, then* $\bigcup_{A \in E} \texttt{Prem}(A)$ *is a preferred subtheory of* Γ*.*

Note that although the above theorem assumes the attack definition of conflict free, it immediately follows from Proposition 16 that Theorem 34 holds if the defeat definition of conflict free is assumed. Finally, also note that the above theorem paves the way for applying argument-game proof theories and labelling algorithms for the stable semantics [34], to preferred subtheories, as well as studying the preferred subtheories approach under the full range of semantics defined for Dung frameworks.

6. A discussion of some related work

6.1. Comparison with general frameworks for argumentation

In this section we compare *ASPIC*$^+$ to related work. To start with, the inclusion of defeasible rules in *ASPIC*$^+$ requires some explanation, given that much current work formalises the construction of arguments as deductive [2,3], and in particular classical [13,26] inference. These approaches regard argumentation-based inference as a form of inconsistency handling in deductive logic; the supposed advantage being that the logic of deductive inference is well-understood [13, p. 16]. This raises the question of whether defeasible inference rules are needed at all. Our answer is that the research history in our field shows that at best only part of argumentation can be formalised as inconsistency handling in deductive logic. To start with, the distinction between strict and defeasible inference rules has a long history in AI research on argumentation [30, 31,37–39,44,46,48], so a truly general framework for structured argumentation must include this distinction. Pollock in particular provides philosophical arguments that appeal to epistemological accounts of human reasoning, so that the modelling of defeasible rules is a particularly salient requirement in light of the bridging role (discussed in Sections 1 and 2) that argumentation plays between human and formal logic-based models of reasoning.

Moreover, conceptually, defeasible reasoning is not about handling inconsistent information but about making deductively unsound but still rational 'jumps' to conclusions on the basis of consistent but deductively inconclusive information. Consider the following well-known example, with the given information that quakers are normally pacifists, that republicans are normally not pacifists and that Richard Nixon was both a quaker and a republican. A defeasible reasoner is then interested in what can be concluded about whether Nixon was a pacifist *while consistently accepting all the given information*. The reason that they are jointly consistent is that 'If q then normally p' and 'q' does not deductively imply p since things could be abnormal: Nixon could be an abnormal quaker (or republican). A defeasible reasoner therefore does not want to reject any of the above statements, but rather wants to assume whenever possible that things are normal, in order to jump to conclusions about Nixon in the absence of evidence to the contrary. In other words, defeasible reasoning is not about inconsistency handling but about making uncertain inferences from consistent (though deductively inconclusive)

premises. Therefore, attempts to formalise defeasible reasoning as inconsistency handling are at least unnatural. Moreover, the literature on non-monotonic logic suggests that such attempts[15] are prone to validating counterintuitive inferences (see e.g. [17,24] or [42] for a recent discussion in the context of argumentation). We therefore conclude that, given the research literature, it makes sense to include defeasible inferences in models of argumentation, and therefore any account of argumentation that claims to be general should leave room for them.

A number of works have been proposed as general approaches to argumentation. A well-known and established framework is that of assumption-based argumentation (*ABA*) [15], which has made a substantial contribution to our understanding of argumentation, and is shown (in [40]) to be a special case of the *ASPIC*$^+$ framework in which arguments are built from assumption premises and strict inference rules only and in which all arguments are equally strong. As mentioned earlier, when commenting on Definition 4, [40]'s result on the relation between *ASPIC*$^+$ and *ABA* also holds if all *ABA* assumptions are translated as *ASPIC*$^+$'s ordinary premises. To see why, firstly, note that *ABA* does not accommodate preferences over assumptions. Hence all undermining attacks on assumption premises are preference independent. Since the reconstruction of *ABA* does not accommodate preferences, then undermining attacks on *ABA* assumptions modelled in this paper as ordinary premises, also always succeed as defeats. One can thus straightforwardly replace [40]'s assumption premises with ordinary premises, and show (as in [40]) that *ABA* can be faithfully reconstructed in this paper's formalisation of *ASPIC*$^+$. Our work is relevant for *ABA*, since *ABA* does not in general satisfy [18]'s consistency postulates.[16] A simple counterexample is an *ABA* deductive system with two rules $\rightarrow p$ and $\rightarrow \neg p$. Note that is not to suggest that *ABA* is flawed; rather, we provide conditions (e.g. that rules be closed under transposition) under which *ABA* satisfies [18]'s consistency postulates.

More recently, Amgoud & Besnard [2,3] proposed the abstract-logic approach (AL) to defining structured argumentation. *ASPIC*$^+$ is considerably more complex than AL: firstly because *ASPIC*$^+$ models the use of preferences to resolve attacks, so it has to distinguish between attack and defeat, and secondly because *ASPIC*$^+$ combines deductive and defeasible argumentation, which means that not only the premises, but also the defeasible inferences of an argument can be attacked. This requires that the arguments' structure be made explicit in order to know which parts of an argument can be attacked. By contrast, if all inferences are certain, then arguments can only be attacked on their premises so their internal structure is irrelevant for their evaluation. In Section 5.2 we straightforwardly extended AL with preferences and then combined the extended AL with *ASPIC*$^+$. However, AL cannot accommodate defeasible inferential rules. This is because the inferential reasoning from premises to conclusion is not rendered explicit, but rather is encoded in AL's single consequence operator, which cannot distinguish between strict and defeasible inference rules: there is no way to distinguish p's and S's for which $p \in Cn(S)$ implies that $S \rightarrow p$ should be in $\mathcal{R}_s$ or $S \Rightarrow p$ should be in $\mathcal{R}_d$. Furthermore, recall that in Section 5.2 we argued that AL's subset minimality condition on premises is not appropriate when accounting for defeasible inference rules.

The inappropriateness of accommodating defeasible argumentation in AL is further illustrated when considering whether *ASPIC*$^+$'s notion of an argument *generates* an abstract logic. If this is the case for a given instance of *ASPIC*$^+$, then all of [2,3]'s results hold for this instance. Suppose an *ASPIC*$^+$ AT and Cn defined as follows[17]:

(1) $p \in Cn(X)$ iff there exists an *ASPIC*$^+$ argument A, with $\texttt{Conc}(A) = p$ and $\texttt{Prem}(A) = X$.

It can be shown that conditions (1), (2) and (3) in the definition of an abstract logic (Definition 23) are satisfied. However (4) is in general not satisfied. Consider an AT with $\mathcal{K} = \{p\}$, $R_s = \emptyset$ and $R_d = \{p \Rightarrow q\}$. Also, (5) is not in general satisfied. Consider any AT with $\mathcal{K} = \emptyset$ and $R_d = \{\Rightarrow p \mid p \in \mathcal{L}\}$.

Of course, many instances of *ASPIC*$^+$ will satisfy (4) and (5), and thus generate abstract logics. But then we should interpret [2,3]'s results, as they apply to these instances, with care. In particular, the notion of consistency of an abstract logic behaves in an unexpected way. Recall that Amgoud & Besnard investigate whether for any Dung-extension E, the set $\bigcup_{(X,p)\in E} X$ is AL consistent (see end of Section 5.2). Now, consider an *ASPIC*$^+$ AT formalising the above Nixon example in a language $\mathcal{L}$ including atoms p, r and q respectively denoting 'Nixon is a pacifist', 'Nixon is a republican' and 'Nixon is a quaker' and a connective $\rightsquigarrow$ for default conditionals. Informally, $\varphi \rightsquigarrow \psi$ means 'if φ then normally ψ'. Let the $^-$ relation correspond to classical negation, $\mathcal{R}_s$ contain all propositionally valid inferences (including $p, \neg p \rightarrow \varphi$ for any $\varphi \in \mathcal{L}$) and $\mathcal{R}_d$ contain a defeasible modus ponens scheme $\varphi, \varphi \rightsquigarrow \psi \Rightarrow \psi$. Then if $\mathcal{K} = \{q, r, q \rightsquigarrow p, r \rightsquigarrow \neg p\}$, any Dung-extension contains all elements of $\mathcal{K}$ as arguments but does not contain arguments for both p and $\neg p$ so any such extension satisfies *indirect consistency* (the closure under strict rules is consistent). However, in the abstract logic generated by Eq. (1), the set $\{q, r, q \rightsquigarrow p, r \rightsquigarrow \neg p\}$ is AL inconsistent, since there exists an *ASPIC*$^+$ argument for every φ by combining the defeasible arguments for p and $\neg p$ (even though this argument is not in any extension).

This discrepancy is caused by the fact that an abstract logic's consequence operator cannot distinguish between strict and defeasible inferences, and so regards a set S as inconsistent if the closure of S under *both* strict and defeasible rules is directly inconsistent. But this consistency requirement is too strong, since the very idea of defeasible reasoning is that one's knowledge need *not* be closed under defeasible inference, since defeasible inference rules can be defeated even if all their antecedents hold.

[15] Including those that make use of applicability predicates to simulate the effects of priorities/preferences.

[16] Just as *ASPIC*$^+$ does not in general satisfy [18]'s postulates, since one is free to instantiate *ASPIC*$^+$ in ways that are not 'well-defined' (Definition 12).

[17] Since for AL, consistency requirements on arguments are added on top of a *given* consequence notion Cn, we cannot incorporate consistency requirements into the *definition* of Cn, and so assume *ASPIC*$^+$ without the restriction to c-consistent arguments.

Concluding our comparison, the abstract logic approach provides a very interesting and insightful generalisation of earlier work on classical argumentation, but does not apply to mixed strict and defeasible argumentation, such as modelled in *ASPIC*$^+$ and earlier by many others. We should here emphasise our view that deductive approaches certainly do have their place in the study and application of argumentation. However, we argue that a truly general account of argumentation should also accommodate the use of defeasible inference rules.

Amgoud & Besnard [2,3] also use the abstract logic approach to make some informal negative claims about the suitability of Dung-style semantics. First of all, they informally claim [3] that to satisfy the consistency postulates, an attack relation should be *valid* in the sense that when two arguments have jointly AL inconsistent premises, they should attack each other. However, this informal claim should be read with care: what they formally show is that validity is a *sufficient* condition for consistency. Their results do not preclude "invalid" attack relations, such as undermining attacks, from satisfying consistency. Indeed, in [40] and this paper we have identified alternative sufficient conditions for consistency. Furthermore, Amgoud & Besnard themselves show that, under the assumption that *a Dung framework contains all arguments that can be logically constructed*, their notion of consistency of extensions is satisfied assuming the AL undermining attacks in Definition 24, which is, of course, consistent with our more general result showing that AL satisfies all of [18] postulates (Remark 31 and Proposition 32 in Section 5.2). Amgoud & Besnard regard this assumption as problematic. However, we regard this not as a problem of the attack relation, but of the reasoner: if an imperfect reasoner is modelled who cannot be relied on to produce all relevant arguments, then perfect results cannot be expected. Furthermore, as argued in Section 2.2, requiring that attacks be 'valid' goes against the dialectical role of attacks and has the computational problem in that it can give rise to infinitely many attacks.

Finally, in a recent publication [1], Amgoud claims that the *ASPIC*$^+$ framework suffers from a number of weaknesses. Space limitations preclude a detailed assessment of these claims here, suffice it to say that the formal results in [40] and in this paper, contradict a number of informal claims in [1]. Furthermore, the interested reader may consult a comprehensive rebuttal of [1]'s claims in [43].

6.2. Comparison with other works on preference-based argumentation

We now consider approaches that accommodate preferences to determine which attacks succeed as defeats. Recently, both Kaci [28] and Amgoud & Vesic [6–9] have addressed the issue of how consistency can be ensured for instantiations of the preference-based argumentation frameworks (*PAF*s) [4] reviewed in Section 2.1. They all argue that instantiations of standard *PAF*s have problems with unsuccessful asymmetric attacks. [28] argues that all attacks should therefore be symmetric. However, [2] shows that for classical argumentation this would still lead to inconsistency problems. Nevertheless, [6,7,9] also criticise 'standard' *PAF* approaches, arguing that unsuccessful asymmetric attacks may violate consistency. As a solution they propose that unsuccessful asymmetric attacks should result in rejection of the attacker even if it is not attacked by any argument. However, our consistency results obviate the need for reversing unsuccessful attacks. We have shown that by taking into account the structure of arguments, one can show that if A unsuccessfully attacks B, then either some sub-argument of B defeats A, or under assumptions on the preference ordering, B can be continued into an argument that defeats A, and that this result is key for showing consistency as discussed in Section 4.3.

But how do our consistency results square with [6,7,9]'s examples of inconsistent *PAF*s? [7] gives a semiformal example which we described earlier in Section 4.3 (and which is also described in terms of uninstantiated abstract arguments in [9]). Recall that an expert's argument A, that a given violin is a Stradivarius (s) and therefore expensive (e), is asymmetrically attacked by a child's argument B that it is not a Stradivarius ($\neg s$). The greater reliability of the expert's assertion about the violin means that A is preferred to B so that B does not defeat A. We observed that inconsistency is not violated under this paper's attack definition of conflict free, since $\{A, B\}$ is not conflict free and so not admissible. However, even under the defeat definition, we can see that [7]'s suggested problem arises only when failing to take into account *all* arguments. Formalising the example in *ASPIC*$^+$, $A = [s; s \Rightarrow e]$ where s is an ordinary premise, and $A' = [s]$ is a sub-argument of A. $B = [\neg s]$ where $\neg s$ is an ordinary premise. Hence B attacks A on A', and the expert's greater reliability means that A' is preferred to B and so B does not defeat A. However, one must also then acknowledge that A' rebut attacks and defeats B, so that $\{A, B\}$ is not admissible.

In [6,9], Amgoud & Vesic give a formal classical logic instantiation of a *PAF* that demonstrates inconsistency. The example used is that formalised here in Example 33. However, Amgoud & Vesic state that A_1 is strictly preferred to the other arguments, all of which are equally preferred. They thus obtain the defeat graph shown in Fig. 5(a), and so the single stable extension $\{A_1, A_2, A_3, A_5\}$, which violates consistency. The difference in outcome arises because [6]'s use of the premise ordering to resolve attacks (which is taken from [4]), differs from our Definition 9 in which if A undermines B on premise p, then A defeats B if $A \not\prec p$. However, in [6], A defeats B if $A \not\prec B$ based on a comparison of *all* premises of B. This makes a crucial difference. Since both A_4 and A_5 have $\neg y$ as weakest premise, A_4 and A_5 are equally preferred in [6]. However, we have that A_5's attack on A_4 is on A_4's sub-argument A_1, so the comparison is between A_5 and A_1. Now since $x >' \neg y$, we have that A_1 is strictly preferred to A_5, so A_5 does not defeat A_4, so A_4 strictly defeats A_5. But then a set including $\{A_1, A_2, A_3, A_5\}$ is not a stable extension, since it does not defend A_5 against A_4. Instead, E'_1 containing the arguments A_1, A_4, and A_2 is stable and satisfies consistency. We prefer our approach over [5,6], since we do not see why the preference of the premise $\neg y$ of A_4, which is irrelevant to the conflict on the premise x, should be relevant in resolving this conflict. Note here that the crucial point is that the structure of arguments and the nature of attack should be taken

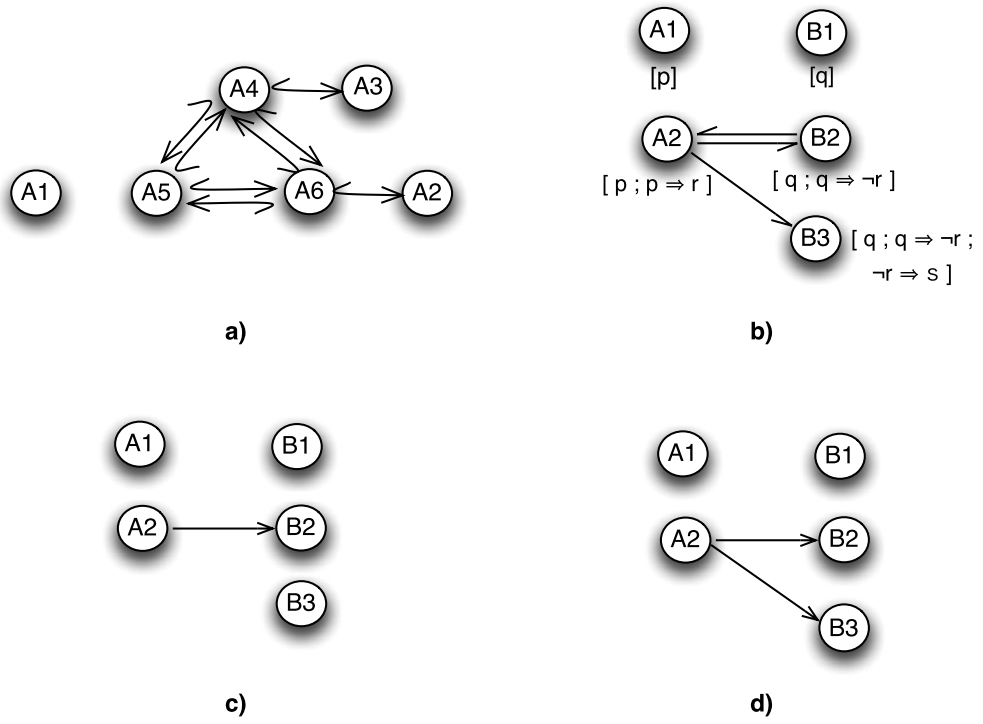

Fig. 5. Examples illustrating comparison with critiques of *PAF*s.

into account when applying preferences. In this case it is crucial to see that A_5's attack on A_4 was a direct attack on A_4's sub-argument A_1.

The issue also arises in different ways. Consider the *ASPIC*$^+$ example, with $\mathcal{K}_p = \{p, q\}$, $\mathcal{K}_n = \emptyset$, $\mathcal{R}_s = \emptyset$, $\mathcal{R}_d = \{p \Rightarrow r; q \Rightarrow \neg r; \neg r \Rightarrow s\}$. We then have the arguments and attacks in Fig. 5(d). Then, assuming $p \Rightarrow r > q \Rightarrow \neg r$ and $\neg r \Rightarrow s > p \Rightarrow r$, the argument ordering $B_2 \prec A_2$, $A_2 \prec B_3$ is generated by the last link principle. A *PAF* modelling then generates the defeat graph in Fig. 5(e), so obtaining the single extension (in whatever semantics) $\{A_1, B_1, A_2, B_3\}$. So not only A_2 but also B_3 is justified. However, not only are A_2 and B_3 based on arguments with contradictory conclusions, but the sub-argument closure postulate is violated; B_3 is justified, but its sub-argument B_2 is not. The problem arises because the *PAF* modelling cannot recognise that A_2 attacks B_3 on its sub-argument B_2, so we should compare A_2 with B_2, and not B_3. Now since $B_2 \prec A_2$, then A_2 defeats B_3, so the single extension (in whatever semantics) is $\{A_1, B_1, A_2\}$ and we have that A_2 is justified and both B_2 and B_3 are overruled, as visualised in Fig. 5(e). Note that these problems are not due to the use of defeasible rules or the last-link ordering. Consider a classical logic instantiation of *ASPIC*$^+$ in which $\mathcal{K}_n = \emptyset$, $\mathcal{K}_p = \{p, q, \neg p\}$ and $q >' \neg p >' p$. The following arguments can be constructed:

$$A_1 = p; \quad A_2 = q, \quad A_3 = p, \quad q \rightarrow p \wedge q \quad \text{and} \quad B = \neg p.$$

Then, A_1 and B attack each other and B attacks A_3 (on p). Suppose arguments are compared based on the weakest link principle, applying Section 5.1's democratic principle to the premise sets. Then $A_1 \prec B$ and $B \prec A_3$. The *PAF* for this example then generates a stable extension containing A_3 and B, which again violates sub-argument closure. In *ASPIC*$^+$ we instead obtain that B defeats A_3 on A_1, so the correct outcome is obtained.

Concluding, [6,7,9] are right that *PAF*s need to be repaired, but the proper repair is not to change definitions at the abstract level but to make the structure of arguments and the nature of attack explicit. We have seen that seeming problems with unsuccessful asymmetric attack at the abstract level disappear if the structure of arguments and the nature of attack are specified, and that seeming violations of postulates do not occur if the success of an attack on an argument X is based on a preference-based comparison on the sub-argument of X that is attacked. We have also seen in this paper that there are reasonable notions of attack that result in defeat irrespective of preferences, such as *ASPIC*$^+$'s undercutting and contrary attacks. A framework that does not make the structure of arguments explicit cannot distinguish between preference dependent and independent attacks.

Finally, note that besides reversing asymmetric attacks, Amgoud & Vesic [6–9] also propose a solution to the problematic cases they identify, and that we have countered above, by using the preference ordering over arguments to define an ordering over sets of arguments, privileging those that are conflict free under the *attack* relation. However this precludes the dialectical use of preferences in deciding the success of attacks between *individual* arguments (as described in Section 2); it is not clear how their use of preferences can be accounted for in dialogues and proof theoretic argument games.

Furthermore, they do not show satisfaction of [18]'s postulates, except in the case of stable extensions, where they show that consistency is satisfied, via a correspondence with Brewka's preferred subtheories [16]. However, we have shown this correspondence without reversing asymmetric attacks, or applying preferences over sets of arguments.

7. Conclusions

A newcomer to the area of abstract argumentation theory might legitimately question its added value above and beyond the conceptual insights yielded by its uniform characterisation of the inference relations of non-monotonic formalisms. This paper began with a response to this rhetorical question. Argumentative characterisations of inference encapsulate the dynamic and dialectical processes of reasoning familiar in everyday debate and discourse.[18] It thus serves to both bridge formal logic and human reasoning in order that the one can inform the other, and support communicative interactions in which heterogeneous agents jointly reason and infer in the presence of uncertainty and conflict. We then discussed the declarative and procedural roles that attacks, preferences and defeats should play in the context of this value proposition, and then reviewed and modified [40]'s *ASPIC*$^+$ framework in light of this discussion. Specifically, the attack relation's denotation of the mutual incompatibility of information in arguments determines whether a given set of arguments is conflict free, as distinct from their possibly preference dependent dialectical use as defeats.

ASPIC$^+$ provides an account of argumentation that combines Dung's argumentation theory with structured arguments, attacks and the use of preferences. The added structure accommodates a range of concrete instantiating logics, to the extent that one can meaningfully study satisfaction of rationality postulates. While the account retains the dialectical apparatus of Dung's theory, one must additionally show that *ASPIC*$^+$'s intermediate level of abstraction allows for a broad range of instantiations, if one is to continue to appeal to the above stated value proposition of argumentation. To this end, we have argued that any general account should accommodate both the traditional use of defeasible inference rules as well as deductive approaches that essentially model non-monotonicity as inconsistency handling. [40]'s version of *ASPIC*$^+$ reconstructed approaches that use defeasible inference rules (e.g., [39,44]) and showed that assumption-based argumentation [15] and systems using argument schemes can be formalised in *ASPIC*$^+$. The modelling of defeasible rules inevitably introduced a degree of complexity that exceeds that of other proposals for general frameworks. However we have argued that a truly general framework for structured argumentation must include defeasible rules. In this paper we adapted *ASPIC*$^+$ to additionally accommodate deductive approaches that require arguments to have consistent premises, and then showed that the adapted *ASPIC*$^+$, with the revised definition of conflict free, satisfies key properties of Dung frameworks and [18]'s rationality postulates under some assumptions. We then formalised instantiation of the adapted *ASPIC*$^+$ with Tarskian (in particular classical) logics extended with preferences, thus demonstrating satisfaction of rationality postulates by these instantiations, and paving the way for the study of other non-classical Tarskian approaches to argumentation. We also addressed some limitations of the way in which argument orderings are defined in [40], and considered a broader range of instantiations of these preference orderings, showing that they satisfy assumptions required for proof of the aforementioned properties and postulates.

Finally, a key rhetorical claim of this paper is that a proper modelling of the use of preferences requires that we take into account the structure of arguments. We believe this claim to be supported by the results in this paper and our discussion of recent critiques of Dung and preference-based argumentation frameworks.

We conclude by mentioning future research. Firstly, we emphasise that *ASPIC*$^+$ is not a system but a framework for specifying systems, such that these systems can be analysed on their properties, for instance, on whether they satisfy the four rationality postulates. An immediate task is to thus show how a range of systems, other than those considered here and in [40,25], can be specified in *ASPIC*$^+$. Secondly, we are currently developing a structured *ASPIC*$^+$ approach to extended argumentation [33], building on a preliminary such structuring in [35]. Thirdly, [19] recently proposed the additional so-called 'non-interference' and 'crash resistance' rationality postulates, which are about whether self-defeating arguments can interfere with the justification status of other arguments in undesired ways. We plan to study the conditions under which these postulates are satisfied by the *ASPIC*$^+$ framework (these have recently been studied [50] in the context of a simplified version of the ASPIC framework [18]). Fourthly, we have in this paper focused on weakest and last link definitions of preference orderings over arguments. We aim in future work to consider other ways of ranking augments, and to study whether such preference orderings satisfy the assumptions identified in this paper for ensuring satisfaction of properties and postulates. Finally, since many conceptual choices made in formalising *ASPIC*$^+$ appeal to the use of argumentation in practice, further real-world applications of *ASPIC*$^+$ are required to establish the framework's utility. One such existing application concerns the use of *ASPIC*$^+$ in modelling the reasoning in a well known legal case [41]. Furthermore, connections between *ASPIC*$^+$ and more informal 'human' modes of argumentative practice need to be established. [14] represents an important first step in this direction, in which *ASPIC*$^+$ is used to provide formal logical foundations for the Argument Interchange Format [22]; an emerging standard for representing argumentation knowledge in both computational and human centred argumentation applications.

[18] Indeed, recent empirically validated work in cognitive science and psychology claims that the cognitive capacity for human reasoning evolved primarily in order to assess and counter the claims and arguments of interlocutors in social settings [32].

Acknowledgements

We would like to thank the anonymous reviewers, whose comments on earlier versions of this paper have helped to improve the content and presentation of this paper.

Appendix A

A.1. Proofs for Section 4.1

Proposition 8. *Let A and B be arguments where B is plausible or defeasible and A and B have contradictory conclusions, and assume $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent if A and B are defined as in Definition 7. Then*:

1. *For all $B' \in M(B)$, there exists a strict continuation $A^+_{B'}$ of $(M(B)\backslash\{B'\}) \cup M(A)$ such that $A^+_{B'}$ rebuts or undermines B on B'.*
2. *If $B \prec A$, and $\preccurlyeq$ is reasonable, then for some $B' \in M(B)$, $A^+_{B'}$ defeats B.*

Proof. 1) Consider first systems closed under contraposition (Definition 12). Observe first that $\texttt{Conc}(M(B)) \cup \texttt{Prem}_\texttt{n}(B) \vdash \texttt{Conc}(B)$ (i.e., one can construct a strict argument concluding $\texttt{Conc}(B)$ with all premises taken from $\texttt{Conc}(M(B))$ and the axiom premises in B). By contraposition, and since $\texttt{Conc}(A)$ and $\texttt{Conc}(B)$ contradict each other, we have that for any $B_i \in M(B)$: $\texttt{Conc}(M(B) \setminus \{B_i\}) \cup \texttt{Prem}_\texttt{n}(B) \cup \texttt{Conc}(A) \vdash -\texttt{Conc}(B_i)$. Hence, one can construct a strict continuation $A^+_{B_i}$ that continues $\{A\} \cup M(B) \setminus \{B_i\} \cup \texttt{Prem}_\texttt{n}(B)$ with strict rules, and that concludes $-\texttt{Conc}(B_i)$.

By construction, $M(B) \setminus \{B_i\}$ and $M(A)$ are the maximal fallible sub-arguments of $A^+_{B_i}$, and $\texttt{Prem}(A^+_{B_i}) \subseteq \texttt{Prem}(A) \cup \texttt{Prem}(B)$.

Since by construction of $M(B)$ either B_i is an ordinary premise or ends with a defeasible inference, $A^+_{B_i}$ either undermines or rebuts B_i. But then $A^+_{B_i}$ also undermines or rebuts B.

For systems closed under transposition the existence of arguments $A^+_{B_i}$ and B_i, for all $B_i \in M(B)$, is proven by straightforward generalisation of Lemma 6 in [18]. Then the proof can be completed as above.

In the case that A and B are defined as in Definition 7, one only need additionally to show that $\texttt{Prem}(A^+_{B_i})$ is c-consistent, which follows given $\texttt{Prem}(A^+_{B_i}) \subseteq \texttt{Prem}(A) \cup \texttt{Prem}(B)$, and $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent by assumption.

2) By construction, each B'continuation $A^+_{B'}$ of A is a strict continuation of $\{A\} \cup M(B) \setminus \{B'\} \cup \texttt{Prem}_\texttt{n}(B)$. Hence, letting $M(B) = \bigcup_{i=1}^{n} B_i$, we have $\{B_1, \ldots, B_n, A\}$ where each $A^+_{B_i}$ is a strict continuation of $\{B_1, \ldots, B_{i-1}, B_{i+1}, B_n, A\}$. Also, B is a strict continuation of $\{B_1, \ldots, B_n\}$. Since $\preccurlyeq$ is reasonable, then by Definition 18-(2), it cannot be that: $B \prec A$ *and* $A^+_{B_1} \prec B_1$ *and* $\ldots$ *and* $A^+_{B_n} \prec B_n$. Since by assumption $B \prec A$, then for some i, $A^+_{B_i}$ rebuts or undermines B on B_i, $A^+_{B_i} \not\prec B_i$, and so $A^+_{B_i}$ defeats B. □

In what follows, recall Notation 5, in which $X \rightharpoonup Y$ denotes X *attacks* Y and $X \hookrightarrow Y$ denotes X *defeats* Y.

Lemma 35. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a (c-)SAF*:

1. *If A is acceptable w.r.t. $S \subseteq \mathcal{A}$ then A is acceptable w.r.t. any superset of S.*
2. *If $A \hookrightarrow B$, then $A \hookrightarrow B'$ for some $B' \in \texttt{Sub}(B)$, and if $A \hookrightarrow B'$, $B' \in \texttt{Sub}(B)$, then $A \hookrightarrow B$.*
3. *If A is acceptable w.r.t. $S \subseteq \mathcal{A}$, $A' \in \texttt{Sub}(A)$, then A' is acceptable w.r.t. S.*

Proof. Proofs of 35-1 and 35-2 are straightforward given the definitions of acceptability and defeat. For 35-3, suppose $B \hookrightarrow A'$. By 35-2, $B \hookrightarrow A$, and so $\exists C \in S$ s.t. $C \hookrightarrow A$. Hence A' is acceptable w.r.t. S. □

Lemma 36. *Suppose $B \rightharpoonup A$, where B attacks A on A', and if A and B are defined as in Definition 7, then $\texttt{Prem}(A) \cup \texttt{Prem}(B)$ is c-consistent. If $B \not\hookrightarrow A$ then either*:

1. *$A' \hookrightarrow B$, or;*
2. *For some $B' \in M(B)$, there is a strict continuation $A'^+_{B'}$ of $(M(B)\backslash\{B'\}) \cup M(A')$ s.t. $A'^+_{B'} \hookrightarrow B$.*

Proof. Since $B \not\hookrightarrow A$, then: B rebuts on the conclusion φ of A' where A''s top rule is defeasible, or B undermines the ordinary premise $A' = \varphi$, and $B \prec A'$. Also, $\texttt{Conc}(B)$ must be a contradictory of φ since otherwise $\texttt{Conc}(B)$ would be a contrary of φ implying that $B \hookrightarrow A$ (by virtue of the preference independent attack by contraries).

Also, B must be plausible or defeasible since for $B \prec A'$ to be the case, B cannot be strict and firm (under the assumption that $\preccurlyeq$ is reasonable (Definition 18)).

1) If B is an ordinary premise or has a defeasible top rule, $A' \rightharpoonup B$, and since $B \prec A'$, $A' \hookrightarrow B$.

2) If B has a strict top rule, then by Proposition 8 there exists a strict continuation $A'^+_{B'}$ s.t. $A'^+_{B'} \hookrightarrow B$. □

The following lemma follows from the fact that if B defeats some strict continuation A of $\{A_1, \ldots, A_n\}$ then the defeat must be on some A_i.

Lemma 37. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a (c-)SAF. Let $A \in \mathcal{A}$ be a strict continuation of $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$, and for $i = 1 \ldots n$, A_i is acceptable w.r.t. $E \subseteq \mathcal{A}$. Then A is acceptable w.r.t. E.*

Proof. Let B be any argument s.t. $B \hookrightarrow A$. By Definition 8, B attacks A by undercutting or rebutting on defeasible rules in A or undermining on an ordinary premise in A. Hence, by definition of strict continuations (Definition 17), it must be that $B \rightharpoonup A$ iff $B \rightharpoonup A_i$ for some (possibly more than one) $A_i \in \{A_1, \ldots, A_n\}$. Either:

1) B undercuts or contrary rebuts/undermines some A_i, and so by Definition 9, B defeats A_i, or:
2) B does not undercut or contrary rebut/undermine some A_i. Suppose for all A_i, for all sub-arguments A'_i of A_i s.t. B rebuts or undermines A_i on A'_i, $B \prec A'_i$. This contradicts B defeats A. Hence, for some A_i, B defeats A_i.

We have shown that if B defeats A then B defeats some A_i. By assumption of A_i acceptable w.r.t. E, $\exists C \in E$ s.t. C defeats B. Hence, A is acceptable w.r.t. E. □

For the following proposition, recall that by assumption, any c-*SAF* is well defined and so satisfies c-classicality (Definition 12).

Proposition 9. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a* c-*SAF. If $A_1, \ldots, A_n$ are acceptable w.r.t. some conflict-free $E \subseteq \mathcal{A}$, then $\bigcup_{i=1}^{n} \mathtt{Prem}(A_i)$ is c-consistent.*

Proof. Suppose for contradiction otherwise, and let S be any minimally c-inconsistent subset of $\bigcup_{i=1}^{n} \mathtt{Prem}(A_i)$. By assumption of c-classicality:

for all $\varphi \in S$, $S \setminus \{\varphi\} \vdash -\varphi$ and $S \setminus \{\varphi\}$ is c-consistent.

We thus have the set of ordinary premises $S = \{\varphi_1, \ldots, \varphi_m\} \subseteq \bigcup_{i=1}^{n} \mathtt{Prem}(A_i)$ (that must be non-empty given that $\mathcal{K}_n$ is c-consistent by assumption of axiom consistency (Definition 12)), such that for $i = 1 \ldots m$, there is a strict continuation $B^{+\setminus i}$ of $\{\varphi_1, \ldots, \varphi_{i-1}, \varphi_{i+1}, \varphi_m\}$ s.t. $B^{+\setminus i} \rightharpoonup \varphi_i$ (recall that elements from $\mathcal{K}$ are also arguments, so this notation is well-defined).

Since $\preccurlyeq$ is reasonable, for some i, $B^{+\setminus i} \not\prec \varphi_i$ and so $B^{+\setminus i} \hookrightarrow \varphi_i$.

Since for $i = 1 \ldots n$, A_i is acceptable w.r.t. E, then: since $\varphi_i \in \bigcup_{i=1}^{n} \mathtt{Prem}(A_i)$, then by Lemma 35-3, φ_i is acceptable w.r.t. E.

Since $B^{+\setminus i}$ is a strict continuation of some subset of $\bigcup_{i=1}^{n} \mathtt{Prem}(A_i)$, then by Lemmas 35-3 and 37, $B^{+\setminus i}$ is acceptable w.r.t. E.

But then since $B^{+\setminus i} \hookrightarrow \varphi_i$, $\exists X, Y \in E$ s.t. $Y \hookrightarrow B^{+\setminus i}$, $X \hookrightarrow Y$, contradicting E is conflict free. □

Lemma 38. *Let A be acceptable w.r.t. an admissible extension S of a (c-)SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$. Then $\forall B \in S \cup \{A\}$, neither $A \hookrightarrow B$ or $B \hookrightarrow A$.*

Proof. Suppose for contradiction that: 1) $A \hookrightarrow B$, $B \in S \cup \{A\}$. By assumption of B's acceptability, $\exists C \in S$ s.t. $C \hookrightarrow A$, and by acceptability of A, $\exists D \in S$ s.t. $D \hookrightarrow C$, hence $D \rightharpoonup C$, contradicting S is conflict free; 2) $B \hookrightarrow A$, $B \in S$. By acceptability of A, $\exists D \in S$ s.t. $D \hookrightarrow B$, hence $D \rightharpoonup B$, contradicting S is conflict free. □

Proposition 10. *Let A be acceptable w.r.t. an admissible extension S of a (c-)SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$. Then $S' = S \cup \{A\}$ is conflict free.*

Proof. Firstly, since for any $B \in S$, B is acceptable w.r.t. S, then by Proposition 9, $\mathtt{Prem}(A) \cup \mathtt{Prem}(B)$ is c-consistent.

Suppose for contradiction that S' is not conflict free. By assumption, S is conflict free. A cannot attack itself since A must then defeat itself, contradicting Lemma 38. Hence, we have the following two cases:

1) $\exists B \in S$, $B \rightharpoonup A$, and $B \not\hookrightarrow A$ by Lemma 38. By Lemma 36, for some sub-argument A' of A, either:
- **1.1**) A' defeats B, hence (by acceptability of B) $\exists C \in S$ s.t. $C \hookrightarrow A'$, and so (by Lemma 35-2) $C \hookrightarrow A$, contradicting Lemma 38, or;
- **1.2**) $\exists A'^{+}_{B'}$ s.t. $A'^{+}_{B'} \hookrightarrow B$, hence $\exists C \in S$ s.t. $C \hookrightarrow A'^{+}_{B'}$. By construction of $A'^{+}_{B'}$ and Lemma 35-2, it must be that $C \hookrightarrow Z$, $Z \in \mathtt{Sub}(A) \cup \mathtt{Sub}(B)$. Hence (by Lemma 35-2) either $C \hookrightarrow B$, contradicting S is conflict free, or $C \hookrightarrow A$, contradicting Lemma 38.

2) $\exists B \in S$, $A \rightharpoonup B$, and $A \not\hookrightarrow B$ by Lemma 38. By Lemma 36, for some sub-argument B' of B, either:
- **2.1**) B' defeats A, hence (by acceptability of A) $\exists C \in S$ s.t. $C \hookrightarrow B'$ and so (by Lemma 35-2) $C \hookrightarrow B$, hence $C \rightharpoonup B$, contradicting S is conflict free, or;

2.2) $\exists B'^{+}_{A'}$ s.t. $B'^{+}_{A'} \hookrightarrow A$, hence $\exists C \in S$ s.t. $C \hookrightarrow B'^{+}_{A'}$. By construction of $B'^{+}_{A'}$, $C \hookrightarrow Z$, $Z \in \mathtt{Sub}(A) \cup \mathtt{Sub}(B)$, leading to a contradiction as in **1.2**). □

Proposition 11. *Let A, A' be acceptable w.r.t. an admissible extension S of a (c-)SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$. Then*:

1. *$S' = S \cup \{A\}$ is admissible.*
2. *A' is acceptable w.r.t. S'.*

Proof. 1) By Lemma 35-1, all arguments in S' are acceptable w.r.t. S'. By Proposition 10, S' is conflict free. Hence S' is admissible. 2) By Lemma 35-1, A' is acceptable w.r.t. S'. □

A.2. Proofs for Section 4.2

Theorem 12 *(Sub-argument closure). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a (c-)SAF and E a complete extension of Δ. Then for all $A \in E$: if $A' \in \mathtt{Sub}(A)$ then $A' \in E$.*

Proof. A' is acceptable w.r.t. E by Lemma 35-3. $E \cup \{A'\}$ is conflict free by Proposition 10. Hence, since E is complete, $A' \in E$. □

Theorem 13 *(Closure under strict rules). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a (c-)SAF and E a complete extension of Δ. Then $\{\mathtt{Conc}(A) | A \in E\} = Cl_{R_s}(\{\mathtt{Conc}(A) | A \in E\})$.*

Proof. It suffices to show that any strict continuation X of $\{A | A \in E\}$ is in E. By Lemma 37, any such X is acceptable w.r.t. E. By Proposition 10, $E \cup \{X\}$ is conflict free. Hence, since E is complete, $X \in E$. Note that if Δ is a c-*SAF*, Proposition 9 guarantees that X's premises are c-consistent. □

Theorem 14 *(Direct consistency). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a (c-)SAF and E an admissible extension of Δ. Then $\{\mathtt{Conc}(A) | A \in E\}$ is consistent.*

Proof. We show that if $A, B \in E$, $\mathtt{Conc}(A) \in \overline{\mathtt{Conc}(B)}$ (i.e., E is inconsistent (Definition 2)), then this leads to a contradiction:

1. A is firm and strict, and:
   - **1.1** if B is strict and firm, then this contradicts the assumption of *axiom consistency* (Definition 12);
   - **1.2** if B is plausible or defeasible, and **1.2.1** B is an ordinary premise or has a defeasible top rule, then $A \rightharpoonup B$, contradicting E is conflict free, or **1.2.2** B has a strict top rule (see **3** below).
2. A is plausible or defeasible, and:
   - **2.1** if B is strict and firm then under the *well-formed* assumption (Definition 12) $\mathtt{Conc}(A)$ cannot be a contrary of $\mathtt{Conc}(B)$, and so they are a contradictory of each other, and **2.1.1** A is an ordinary premise or has a defeasible top rule, in which case $B \rightharpoonup A$, contradicting E is conflict free, or **2.1.2** A has a strict top rule (see **3** below);
   - **2.2** if B is plausible or defeasible and **2.2.1** B is an ordinary premise or has a defeasible top rule then $A \rightharpoonup B$, contradicting E is conflict free, or **2.2.2** B has a strict top rule (see **3** below).
3. Each of **1.2.2**, **2.1.2** and **2.2.2** describes the case where $X, Y \in E$, $\mathtt{Conc}(X) \in \overline{\mathtt{Conc}(Y)}$, Y is defeasible or plausible and has a strict top rule, and so by the *well-formed* assumption $\mathtt{Conc}(X)$ and $\mathtt{Conc}(Y)$ must be contradictory.
   In the case that Δ is a c-*SAF*, since $X, Y \in E$, then X, Y are acceptable w.r.t. E, and so by Proposition 9, $\mathtt{Prem}(A) \cup \mathtt{Prem}(B)$ is c-consistent.
   By Proposition 8 there is a strict continuation $X^{+}_{Y'}$ of $M(Y) \backslash \{Y'\} \cup M(X)$ s.t. $X^{+}_{Y'} \rightharpoonup Y$. By Lemma 37 $X^{+}_{Y'}$ is acceptable w.r.t. E, and by Proposition 10, $E \cup \{X^{+}_{Y'}\}$ is conflict free, contradicting $X^{+}_{Y'} \rightharpoonup Y$. □

Theorem 15 *(Indirect consistency). Let $\Delta = (\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a (c-)SAF and E a complete extension of Δ. Then $Cl_{\mathcal{R}_s}(\{\mathtt{Conc}(A) | A \in E\})$ is consistent.*

Proof. Follows from Theorems 13 and 14. □

A.3. Proofs for Section 4.3

Proposition 16. *Let Δ be a (c-)SAF. For $T \in$ {admissible, complete, grounded, preferred, stable}, E is an* att-T *extension of Δ iff E is a* def-T *extension of Δ.*

Proof. We first show that E is conflict free under the attack definition iff E is conflict free under the defeat definition. The left to right half is trivial: if no two arguments in E attack each other, then no two arguments in E defeat each other. For the right to left half, suppose $B, A \in E$, $B \rightharpoonup A$, $B \not\prec A$. First note that since A, B are acceptable w.r.t. E, then in the case of a *c-SAF* where A and B are defined as in Definition 7, $\mathtt{Prem}(A) \cup \mathtt{Prem}(B)$ is c-consistent by Proposition 9. Then, by Lemma 36, $\exists A' \in \mathtt{Sub}(A)$ s.t. either: i) $A' \hookrightarrow B$, or ii) there is a strict continuation $A'^{+}_{B'}$ of $(M(B)\backslash\{B'\}) \cup M(A')$ s.t. $A'^{+}_{B'} \hookrightarrow B$. In case i), (by acceptability of B) $\exists C \in E$ s.t. $C \hookrightarrow A'$, and so (by Lemma 35-2) $C \hookrightarrow A$, contradicting E is defeat conflict free. In case ii), (by acceptability of B), $\exists C \in E$ s.t. $C \hookrightarrow A'^{+}_{B'}$. By construction of $A'^{+}_{B'}$ and Lemma 35-2, $C \hookrightarrow Z$, $Z \in \mathtt{Sub}(A) \cup \mathtt{Sub}(B)$. Hence, (by Lemma 35-2) either $C \hookrightarrow B$ or $C \hookrightarrow A$, contradicting E is defeat conflict free.

Next, note that admissible and complete extensions are in Definition 1 defined in terms of conflict-freeness and acceptability, where acceptability is according to Definition 15 defined in terms of defeat relations between arguments. Then since any *att* semantics and *def* semantics agree on the defeat relation between arguments, the proposition follows for admissible and complete semantics. Then since preferred and grounded semantics are defined in terms of complete semantics, it also follows for these semantics, and then since stable semantics is defined in terms of preferred semantics and the defeat relation, it also follows for stable semantics. □

A.4. Proofs for Section 5.1

In the following proofs, we may write $\mathtt{LDR}$ as an abbreviation for $\mathtt{LastDefRules}$, and $\mathtt{DR}$ as an abbreviation for $\mathtt{DefRules}$. Also, as an abuse of notation we may simply write $\trianglelefteq$ instead of $\trianglelefteq_{\mathtt{s}}$.

Proposition 19. *Let $\preccurlyeq$ be defined according to the last-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is reasonable inducing. Then $\preccurlyeq$ is* reasonable.

Proof. *Proof of the first condition of reasonableness*:

1-i) Assume A is strict and firm, and so $\mathtt{LDR}(A) = \emptyset$ and $\mathtt{Prem}_{\mathtt{p}}(A) = \emptyset$.
- If $\mathtt{LDR}(B) \neq \emptyset$, then A and B must be compared by the first condition of Definition 20. By Definition 19, $\mathtt{LDR}(B) \trianglelefteq \mathtt{LDR}(A)$, $\mathtt{LDR}(A) \not\trianglelefteq \mathtt{LDR}(B)$, and so $B \preccurlyeq A$, $A \not\preccurlyeq B$, and so $B \prec A$.
- If $\mathtt{LDR}(B) = \emptyset$, then A and B must be compared by the second condition of Definition 20. By assumption of B being plausible or defeasible, $\mathtt{Prem}_{\mathtt{p}}(B) \neq \emptyset$. By Definition 19, $\mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$, and $\mathtt{Prem}_{\mathtt{p}}(A) \not\trianglelefteq \mathtt{Prem}_{\mathtt{p}}(B)$, and so $B \preccurlyeq A$, $A \not\preccurlyeq B$, and so $B \prec A$.

1-ii) Assume B is strict and firm, and so $\mathtt{LDR}(B) = \emptyset$, $\mathtt{Prem}_{\mathtt{p}}(B) = \emptyset$. Then by Definition 19, $\mathtt{LDR}(B) \not\trianglelefteq \mathtt{LDR}(A)$ and $\mathtt{Prem}_{\mathtt{p}}(B) \not\trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$, and so it cannot be that $B \preccurlyeq A$ and so $B \prec A$, by the first or second conditions of Definition 20.

1-iii) Follows straightforwardly from Definition 20, given that A' differs from A only in its strict rules and/or axiom premises.

Proof of the second condition of reasonableness:

Suppose for contradiction that:

$$\forall i,\ \text{there is a strict continuation } C^{+\backslash i} \text{ of } \{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\} \text{ such that } C^{+\backslash i} \prec C_i. \qquad \textbf{i})$$

1) Suppose for some $i = 1 \ldots n$, $\mathtt{LDR}(C_i) \neq \emptyset$. Then, it $C^{+\backslash i} \preccurlyeq C_i$ since condition 1 of Definition 20 holds, i.e., $\mathtt{LDR}(C^{+\backslash i}) \trianglelefteq \mathtt{LDR}(C_i)$. Since $C^{+\backslash i} \prec C_i$, then $C_i \not\trianglelefteq C^{+\backslash i}$, and so $\mathtt{LDR}(C_i) \not\trianglelefteq \mathtt{LDR}(C^{+\backslash i})$, and so $\bigcup_{j=1, j\neq i}^{n} \mathtt{LDR}(C_j) \triangleleft \mathtt{LDR}(C_i)$.

By Definition19-(1), it must be that $\bigcup_{j=1, j\neq i}^{n} \mathtt{LDR}(C_j) \neq \emptyset$. Let $\{C_j, \ldots, C_m\}$ be the subset of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$ s.t. for $k = j \ldots m$, $\mathtt{LDR}(C_k) \neq \emptyset$. Then:

$$\bigcup_{k=j}^{m} \mathtt{LDR}(C_k) \triangleleft \mathtt{LDR}(C_i). \qquad \textbf{ii})$$

By virtue of $\trianglelefteq$ satisfying Definition 22-(2a) for some $k = j \ldots m$, $\mathtt{LDR}(C_k) \trianglelefteq \mathtt{LDR}(C_i)$. Then, given that by **i**), $C^{+\backslash k} \prec C_k$, and given that $\mathtt{LDR}(C_k) \neq \emptyset$, one can reason in the same way, concluding that for some $l = j \ldots m$, $l \neq k$, $\mathtt{LDR}(C_l) \trianglelefteq \mathtt{LDR}(C_k)$.

Since $\{C_k, \ldots, C_m\}$ is finite, we can continue reasoning in the same way, obtaining that $\mathtt{LDR}(C_i) \trianglelefteq \cdots \trianglelefteq \mathtt{LDR}(C_m) \trianglelefteq \mathtt{LDR}(C_i)$. Then by transitivity of $\trianglelefteq$, for $k = j \ldots m$, $\mathtt{LDR}(C_i) \trianglelefteq \mathtt{LDR}(C_k)$. But given **ii**), then by virtue of $\trianglelefteq$ satisfying Definition 22-(2b), for some $k = j \ldots m$, $\mathtt{LDR}(C_i) \not\trianglelefteq \mathtt{LDR}(C_k)$. Contradiction.

2) Suppose for $i = 1 \ldots n$, $\mathtt{LDR}(C_i) = \emptyset$. Then, given **i**), we can conclude (as above) that $\bigcup_{j=1, j\neq i}^{n} \mathtt{Prem}_{\mathtt{p}}(C_j) \triangleleft \mathtt{Prem}_{\mathtt{p}}(C_i)$. One can then reason as above (by virtue of $\trianglelefteq$ satisfying Definitions 22-(2a) and 22-(2b)), to conclude that $\mathtt{Prem}_{\mathtt{p}}(C_i) \trianglelefteq \cdots \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(C_n) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(C_i)$, leading to a contradiction as above. □

Proposition 20. *Let $\preccurlyeq$ be defined according to the weakest-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is reasonable inducing. Then $\preccurlyeq$ is* reasonable.

Proof. *Proof of the first condition of reasonableness*:

1-i) Assume A is strict and firm, and so $\texttt{LDR}(A) = \emptyset$ and $\texttt{Prem}_\texttt{p}(A) = \emptyset$:
- Suppose B is strict. Then by assumption, B is plausible, i.e., $\texttt{Prem}_\texttt{p}(B) \neq \emptyset$. Hence by Definition 19, $\texttt{Prem}_\texttt{p}(B) \trianglelefteq \texttt{Prem}_\texttt{p}(A)$, $\texttt{Prem}_\texttt{p}(A) \not\trianglelefteq \texttt{Prem}_\texttt{p}(B)$ and so by Definition 21-1), $B \preccurlyeq A$, $A \not\preccurlyeq B$, and so $B \prec A$.
- Suppose B is firm. Then by assumption B is defeasible, i.e., $\texttt{DR}(B) \neq \emptyset$. Hence by Definition 19, $\texttt{DR}(B) \trianglelefteq \texttt{DR}(A)$, $\texttt{DR}(A) \not\trianglelefteq \texttt{DR}(B)$, and so by Definition 21-2), $B \preccurlyeq A$, $A \not\preccurlyeq B$, and so $B \prec A$.
- Suppose B is defeasible and plausible. Then by Definition 19, $\texttt{Prem}_\texttt{p}(B) \trianglelefteq \texttt{Prem}_\texttt{p}(A)$, $\texttt{Prem}_\texttt{p}(A) \not\trianglelefteq \texttt{Prem}_\texttt{p}(B)$, and $\texttt{DR}(B) \trianglelefteq \texttt{DR}(A)$, $\texttt{DR}(A) \not\trianglelefteq \texttt{DR}(B)$, and so by Definition 21-(3), $B \preccurlyeq A$, $A \not\preccurlyeq B$, and so $B \prec A$.

1-ii) Assume B is strict and firm, and so $\texttt{LDR}(B) = \emptyset$, $\texttt{Prem}_\texttt{p}(B) = \emptyset$. Then by Definition 19, it cannot be that $\texttt{LDR}(B) \trianglelefteq \texttt{LDR}(A)$ or $\texttt{Prem}_\texttt{p}(B) \trianglelefteq \texttt{Prem}_\texttt{p}(A)$, and so it cannot be that $B \preccurlyeq A$ and so $B \prec A$, by the first, second or third conditions of Definition 21.

1-iii) Follows straightforwardly from Definition 21, given that A' differs from A only in its strict rules and/or axiom premises.

Proof of the second condition of reasonableness for the weakest link principle:

Suppose for contradiction that:

$$\forall i,\ \text{there is a strict continuation } C^{+\backslash i} \text{ of } \{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\} \text{ such that } C^{+\backslash i} \prec C_i. \qquad \textbf{i})$$

Given 1-i) and 1-ii), at least one argument C_i must be defeasible or plausible.

1) Suppose for some $i = 1 \ldots n$, $\texttt{DR}(C_i) \neq \emptyset$. Then, the assumed weakest link preference $C^{+\backslash i} \prec C_i$ holds on the basis of Definition 21-2) or 21-3). One can then reason in exactly the same way as in the proof of the second condition of reasonableness in Proposition 19 – case 1) – substituting '$\texttt{DR}$' for '$\texttt{LDR}$', showing that this leads to a contradiction.
2) Suppose for some $i = 1 \ldots n$, $\texttt{Prem}_\texttt{p}(C_i) \neq \emptyset$. Then, $C^{+\backslash i} \prec C_i$ by Definition 21-1) or 21-3). One can then reason in exactly the same way as in the proof of the second condition of reasonableness in Proposition 19 – case (1) – substituting '$\texttt{Prem}_\texttt{p}$' for '$\texttt{LDR}$', showing that this leads to a contradiction. □

Proposition 21. $\trianglelefteq_{\texttt{Eli}}$ *is reasonable inducing.*

Proof. We show that $\trianglelefteq_{\texttt{Eli}}$ is transitive:

Suppose $\Gamma \trianglelefteq_{\texttt{Eli}} \Gamma' \trianglelefteq_{\texttt{Eli}} \Gamma''$. By Definition 19 it must be that $\Gamma \neq \emptyset$, $\Gamma' \neq \emptyset$. If $\Gamma'' = \emptyset$ then $\Gamma \trianglelefteq_{\texttt{Eli}} \Gamma''$ by Definition 19. Else if $\Gamma'' \neq \emptyset$, $\exists X \in \Gamma$ s.t. $\forall X' \in \Gamma'$, $X \leqslant X'$, and $\exists X' \in \Gamma'$ s.t. $\forall X'' \in \Gamma''$, $X' \leqslant X''$. Hence by transitivity of $\leqslant$, $\exists X \in \Gamma$ s.t. $\forall X'' \in \Gamma''$, $X \leqslant X''$, i.e., $\Gamma \trianglelefteq_{\texttt{Eli}} \Gamma''$.

We show that $\trianglelefteq_{\texttt{Eli}}$ satisfies the properties in Definition 22(2a) and 22(2b):

Assume $\bigcup_{i=1}^{n} \texttt{kr}(B_i) \lhd_{\texttt{Eli}} \texttt{kr}(A)$.

By Definition 19, it must be that $\bigcup_{i=1}^{n} \texttt{kr}(B_i) \neq \emptyset$. If $\texttt{kr}(A) = \emptyset$ then (1) and (2) are shown by Definition 19-2) and 19-1) respectively.

Suppose $\texttt{kr}(A) \neq \emptyset$. By assumption:

(1) $\exists Y \in \bigcup_{i=1}^{n} \texttt{kr}(B_i)$ s.t. $\forall X \in \texttt{kr}(A)$, $Y \leqslant X$, and so for some $i = 1 \ldots n$, $Y \in \texttt{kr}(B_i)$, and $\forall X \in \texttt{kr}(A)$, $Y \leqslant X$.

(2) Without loss of generality, let us assume:

$$\texttt{kr}(B_1) = \{Y_1, \ldots, Y_n\}, \qquad A = \{X_1, \ldots, X_m\}, \quad \text{and} \quad Y_1 \leqslant X_1, \ldots, Y_1 \leqslant X_m. \qquad \textbf{1})$$

Now, suppose for contradiction that $\texttt{kr}(A) \trianglelefteq \texttt{kr}(B_1) \ldots \texttt{kr}(A) \trianglelefteq \texttt{kr}(B_n)$. Given $\texttt{kr}(A) \trianglelefteq \texttt{kr}(B_1)$, then without loss of generality, assume:

$$X_1 \leqslant Y_1, \ldots, X_1 \leqslant Y_n. \qquad \textbf{2})$$

By assumption, $\texttt{kr}(A) \not\trianglelefteq_{\texttt{Eli}} \bigcup_{i=1}^{n} \texttt{kr}(B_i)$; i.e., it is not the case that $\exists X \in \texttt{kr}(A)$ s.t. $\forall Y \in \bigcup_{i=1}^{n} \texttt{kr}(B_i)$, $X \leqslant Y$. That is:

$$\forall X \in \texttt{kr}(A), \quad \exists Y \in \bigcup_{i=1}^{n} \texttt{kr}(B_i) \text{ s.t. } \neg(X \leqslant Y). \qquad \textbf{3})$$

Hence, given **2**), it must be that

$$\text{for some } Y \in \bigcup_{i=1}^{n} \texttt{kr}(B_i),\ Y \notin \texttt{kr}(B_1),\ X_1 \not\leqslant Y. \qquad \textbf{4})$$

But then given that we have assumed for contradiction that $\texttt{kr}(A) \trianglelefteq \texttt{kr}(B_1) \ldots \texttt{kr}(A) \trianglelefteq \texttt{kr}(B_n)$, then:

$$\forall Y \in \bigcup_{i=1}^{n} \texttt{kr}(B_i), \quad \exists X \in \texttt{kr}(A), \quad X \leqslant Y.$$

And so by **1**) and transitivity of $\leqslant$, $\forall Y \in \bigcup_{i=1}^{n} \mathtt{kr}(B_i)$, $Y_1 \leqslant Y$, and so by **2**) and transitivity of $\leqslant$, $\forall Y \in \bigcup_{i=1}^{n} \mathtt{kr}(B_i)$, $X_1 \leqslant Y$, contradicting **4**). □

Proposition 22. *$\trianglelefteq_{\mathtt{Dem}}$ is reasonable inducing.*

Proof. We show that $\trianglelefteq_{\mathtt{Dem}}$ is transitive:

Suppose $\Gamma \trianglelefteq_{\mathtt{Dem}_1} \Gamma' \trianglelefteq_{\mathtt{Dem}_1} \Gamma''$. By Definition 19-1) it must be that $\Gamma \neq \emptyset$, $\Gamma' \neq \emptyset$.

If $\Gamma'' = \emptyset$ then $\Gamma \trianglelefteq_{\mathtt{Dem}} \Gamma''$ by Definition 19-2). Else if $\Gamma'' \neq \emptyset$, $\forall X \in \Gamma$, $\exists X' \in \Gamma'$ s.t. $X \leqslant X'$, and $\forall X' \in \Gamma'$, $\exists X'' \in \Gamma''$, s.t. $X' \leqslant X''$. Hence, by transitivity of $\leqslant$, $\forall X \in \Gamma$, $\exists X'' \in \Gamma''$ s.t. $X \leqslant X''$, i.e., $\Gamma \trianglelefteq_{\mathtt{Dem}_1} \Gamma''$.

We show that $\trianglelefteq_{\mathtt{Dem}}$ satisfies the properties in Definition 22(2a) and 22(2b):

Assume $\bigcup_{i=1}^{n} \mathtt{kr}(B_i) \lhd_{\mathtt{Dem}} \mathtt{kr}(A)$. By Definition 19-1), it must be that $\bigcup_{i=1}^{n} \mathtt{kr}(B_i) \neq \emptyset$. If $\mathtt{kr}(A) = \emptyset$ then (1) and (2) are shown by Definition 19-2) and 19-1) respectively.

Suppose $\mathtt{kr}(A) \neq \emptyset$. By assumption:

(1) $\forall Y \in \bigcup_{i=1}^{n} \mathtt{kr}(B_i)$, $\exists X \in \mathtt{kr}(A)$, $Y \leqslant X$. Hence for some $i = 1 \ldots n$, $\forall Y \in \mathtt{kr}(B_i)$, $\exists X \in \mathtt{kr}(A)$, $Y \leqslant X$.

(2) By assumption, $\mathtt{kr}(A) \not\trianglelefteq_{\mathtt{Dem}} \bigcup_{i=1}^{n} \mathtt{kr}(B_i)$; i.e., it is not the case that $\forall X \in \mathtt{kr}(A)$, $\exists Y \in \bigcup_{i=1}^{n} \mathtt{kr}(B_i)$, $X \leqslant Y$. That is:

$$\exists X \in \mathtt{kr}(A) \quad \text{s.t.} \quad \forall Y \in \bigcup_{i=1}^{n} \mathtt{kr}(B_i),\ \neg(X \leqslant Y). \tag{5}$$

Suppose for some B_i, $\mathtt{kr}(A) \trianglelefteq \mathtt{kr}(B_i)$. Then $\forall X \in \mathtt{kr}(A)$, $\exists Y \in \mathtt{kr}(B_i)$ s.t. $X \leqslant Y$, contradicting **5**). □

Proposition 23. *Let $\preccurlyeq$ be defined according to the last-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is transitive. Then $\prec$ is a strict partial order.*

Proof. *Irreflexivity* is immediate given that if $A \prec A$, then $A \preccurlyeq A$, $A \not\preccurlyeq A$. Contradiction.

Transitivity: Suppose $C \prec B \prec A$. To show transitivity we show that $C \prec A$, i.e.: 1) $C \preccurlyeq A$, and 2) $A \not\preccurlyeq C$:

1) Since $C \prec B \prec A$ then $C \preccurlyeq B \preccurlyeq A$. We show that $\preccurlyeq$ is transitive, i.e., $C \preccurlyeq A$: Consider the following two cases:
1) Suppose $\mathtt{LDR}(A) \neq \emptyset$. Then $B \preccurlyeq A$ by condition 1 of Definition 20; i.e., $\mathtt{LDR}(B) \trianglelefteq \mathtt{LDR}(A)$, and so by Definition 19-1) it must be that $\mathtt{LDR}(B) \neq \emptyset$. By the same reasoning, $\mathtt{LDR}(C) \trianglelefteq \mathtt{LDR}(B)$, and $\mathtt{LDR}(C) \neq \emptyset$. By transitivity of $\trianglelefteq$, $\mathtt{LDR}(C) \trianglelefteq \mathtt{LDR}(A)$ and so $C \preccurlyeq A$.
2) Suppose $\mathtt{LDR}(A) = \emptyset$.
   – If $\mathtt{LDR}(C) \neq \emptyset$, then $\mathtt{LDR}(C) \trianglelefteq \mathtt{LDR}(A)$ by Definition 19-2), and so $C \preccurlyeq A$ by condition 1 of Definition 20.
   – If $\mathtt{LDR}(C) = \emptyset$ then it must be that $\mathtt{LDR}(B) = \emptyset$ (if $\mathtt{LDR}(B) \neq \emptyset$, then by Definition 19 it cannot be that $\mathtt{LDR}(C) \trianglelefteq \mathtt{LDR}(B)$, and so one could not conclude $C \preccurlyeq B$ by condition 1 or 2 of Definition 20). Hence $B \preccurlyeq A$ and $C \preccurlyeq B$ by condition 2 of Definition 20. That is to say, $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$. By transitivity of $\trianglelefteq$, $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$, and so $C \preccurlyeq A$.

2) Suppose $A \preccurlyeq C$. Then by transitivity, $B \preccurlyeq C$, contradicting $C \prec B$. □

Proposition 24. *Let $\preccurlyeq$ be defined according to the weakest-link principle, based on a set comparison $\trianglelefteq_{\mathtt{s}}$ that is transitive. Then $\prec$ is a strict partial order.*

Proof. *Irreflexivity* is immediate given that if $A \prec A$, then $A \preccurlyeq A$, $A \not\preccurlyeq A$. Contradiction.

Transitivity: Suppose $C \prec B \prec A$. It suffices to show that $\preccurlyeq$ is transitive, i.e., $C \preccurlyeq A$ (since as in the proof of Proposition 23, we can then show that $A \preccurlyeq C$ leads to a contradiction). Consider the following two cases:

1) C and B are strict. Since B is strict then A must be strict, since otherwise, $B \preccurlyeq A$ holds by virtue of Definition 21-(2) or 21-(3), and so $\mathtt{DR}(B) \trianglelefteq \mathtt{DR}(A)$. But then by Definition 19-1) this cannot be the case since $\mathtt{DR}(B) = \emptyset$. Hence $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$, and so by transitivity of $\trianglelefteq$, $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$. Hence $C \preccurlyeq A$ by condition 1 of Definition 21.
2) C and B are firm. Since B is firm then A must be firm, since otherwise, $B \preccurlyeq A$ holds by virtue of Definition 21-(1) or 21-(3), and so $\mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$. But then by Definition 19-1) this cannot be the case since $\mathtt{Prem}_{\mathtt{p}}(B) = \emptyset$. Hence $\mathtt{DR}(C) \trianglelefteq \mathtt{DR}(B) \trianglelefteq \mathtt{DR}(A)$, and so by transitivity of $\trianglelefteq$, $\mathtt{DR}(C) \trianglelefteq \mathtt{DR}(A)$. Hence $C \preccurlyeq A$ by condition 2 of Definition 21.
3) C and B fall into neither of the above two cases. Then $\mathtt{DR}(C) \trianglelefteq \mathtt{DR}(B)$ and $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(B)$. Hence, by Definition 19-1) it must be that $\mathtt{DR}(C) \neq \emptyset$, $\mathtt{Prem}_{\mathtt{p}}(C) \neq \emptyset$.
   – If $\mathtt{DR}(A) = \emptyset$, then by Definition 19-2), $\mathtt{DR}(C) \trianglelefteq \mathtt{DR}(A)$. If $\mathtt{DR}(A) \neq \emptyset$, then given $B \preccurlyeq A$, $\mathtt{DR}(B) \trianglelefteq \mathtt{DR}(A)$, and so by transitivity of $\trianglelefteq$, $\mathtt{DR}(C) \trianglelefteq \mathtt{DR}(A)$.
   – If $\mathtt{Prem}_{\mathtt{p}}(A) = \emptyset$, then by Definition 19-2), $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$. If $\mathtt{Prem}_{\mathtt{p}}(A) \neq \emptyset$, then given $B \preccurlyeq A$, $\mathtt{Prem}_{\mathtt{p}}(B) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$, and so by transitivity of $\trianglelefteq$, $\mathtt{Prem}_{\mathtt{p}}(C) \trianglelefteq \mathtt{Prem}_{\mathtt{p}}(A)$.

   Hence $C \preccurlyeq A$ by condition 3 of Definition 21.

A.5. Proofs for Section 5.2

Lemma 39. *Let $(AS, \mathcal{K})$ be the abstract logic argumentation theory based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Then for all $X \subseteq \Sigma$ it holds that $p \in Cn(X)$ iff $X \vdash p$.*

Proof. From left to right, suppose $p \in Cn(X)$ for some $X \subseteq \Sigma$. By Definition 23-(3), X is finite, so $X \rightarrow p \in \mathcal{R}_s$ (by Definition 25), so $X \vdash p$ (recall that $X \vdash p$ denotes a strict *ASPIC*$^+$ argument for p based on premises $X' \subseteq X$).

From right to left is proven by induction on the structure of arguments. Assume $A = X \vdash p$, where by Definition 25, $X \subseteq \mathcal{K}_p$, $X \subseteq \Sigma$. Assume first $p \in X$. Then $p \in Cn(X)$ by Definition 23-(1). Consider next any $A = A_1, \ldots, A_n \rightarrow \varphi$. By inductive hypothesis, $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \in Cn(X)$. Since $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \varphi \in \mathcal{R}_s$, then by Definition 25, $\varphi \in Cn(\bigcup_{i=1}^{n} Conc(A_i))$. Since $\bigcup_{i=1}^{n} Conc(A_i) \subseteq Cn(X)$, then by monotonicity, $\varphi \in Cn(Cn(X))$. But then by Definition 23-(2), $\varphi \in Cn(X)$. □

Lemma 40. *Let $(\mathcal{L}, Cn)$ be an abstract logic. For any finite $S \subseteq \mathcal{L}$ and any $p \in \mathcal{L}$, if $S \cup \{p\}$ is AL inconsistent, then there exists an $s \in Cn(S)$ such that $\{s, p\}$ is AL-inconsistent.*

Proof. Since S is finite, we can with repeated application of the definition of adjunction conclude that there exists an s such that $Cn(\{s\}) = Cn(S)$. By Definition 23-(1), $s \in Cn(\{s\})$ and so $s \in Cn(S)$. By Definition 23-(2), $Cn(S \cup \{p\}) = Cn(\{s\} \cup \{p\})$. Then $Cn(\{s\} \cup \{p\}) = \mathcal{L}$ so $\{s, p\}$ is AL-inconsistent. □

Lemma 41. *If $X \cup Z$ is AL-inconsistent and $X \subseteq Cn(Y)$ then $Y \cup Z$ is AL-inconsistent.*

Proof. We prove the contraposition that if $Y \cup Z$ is AL-consistent and $X \subseteq Cn(Y)$, then $X \cup Z$ is AL-consistent. To prove this, we first show that:

$$\text{If } X \subseteq Cn(Y), \text{ then } Cn(X \cup Y) = Cn(Y). \tag{1}$$

By monotonicity, $Cn(Y) \subseteq Cn(X \cup Y)$. We prove that $Cn(X \cup Y) \subseteq Cn(Y)$. Let $X \subseteq Cn(Y)$. By Definition 23-(1), $Y \subseteq Cn(Y)$. But then $X \cup Y \subseteq Cn(Y)$. Then by monotonicity $Cn(X \cup Y) \subseteq Cn(Cn(Y))$. Since $Cn(Cn(Y)) = Cn(Y)$ (by Definition 23-(2)), then $Cn(X \cup Y) \subseteq Cn(Y)$. So $Cn(X \cup Y) = Cn(Y)$.

We have shown (**1**). Then by property (7) in Section 5.2, $Cn(X \cup Y \cup Z) = Cn(Y \cup Z)$. Now if $Cn(Y \cup Z) \neq \mathcal{L}$ then $Cn(X \cup Y \cup Z) \neq \mathcal{L}$. But then by monotonicity $Cn(X \cup Z) \neq \mathcal{L}$. □

Proposition 25. *A c-SAF based on an AL argumentation theory is closed under contraposition, axiom consistent, c-classical, and well-formed.*

Proof. Well-formedness immediately follows from the fact that the $^-$ relation is symmetric. Axiom consistency follows from the fact that $\mathcal{K}_n = \emptyset$. To prove satisfaction of contraposition we must prove:

$$\text{If } S \vdash p \text{ then for all } s \in S \text{ it holds that } S \setminus \{s\} \cup \{-p\} \vdash -s \text{ for any } -p \text{ and } -s.$$

By definition of $\vdash$, if $S \vdash p$ then $S' \vdash p$ for some finite $S' \subseteq S$. Therefore we can without loss of generality assume that S is finite. Next, if $S \vdash p$ then $p \in Cn(S)$ by Lemma 39. Consider any $-p$. Then $\{p, -p\}$ is AL-inconsistent by Definition 25(3c). But then $S \cup \{-p\}$ is AL-inconsistent by Lemma 41. Then by simple rewriting for all $s \in S$ it holds that $(S \setminus \{s\}) \cup \{-p\} \cup \{s\}$ is AL-inconsistent. By assumption that S is finite, Lemma 40 then yields that there exists an $s' \in Cn(S \setminus \{s\} \cup \{-p\})$ such that $\{s', s\}$ is AL-inconsistent. Then by Definition 25(3c) there exists an $s'' \in Cn(\{s'\})$ such that $s'' = -s$. By Definition 23-(2) and monotonicity, $s'' \in Cn(S \setminus \{s\} \cup \{-p\})$. Hence, $-s \in Cn(S \setminus \{s\} \cup \{-p\})$. Hence, $S \setminus \{s\} \cup \{-p\} \vdash -s$ by Lemma 39.

C-classicality is proven as follows. We first prove that if $S \subseteq \mathcal{L}$ is AL-inconsistent, then some finite $S' \subseteq S$ is AL-inconsistent. By Definition 23-(4), $\exists p \in \mathcal{L}$ such that $Cn(p) = \mathcal{L}$. Let p be denoted by $\bot$. Now suppose S is inconsistent. Then $Cn(S) = \mathcal{L}$, so $\bot \in Cn(S)$. By Definition 23-(3), $\bot \in Cn(S')$ for some finite $S' \subseteq S$. But since $Cn(Cn(S')) = Cn(S')$ (Definition 23-(2)), this implies that $Cn(S') = \mathcal{L}$.

Now assume $S \subseteq \mathcal{L}$ is minimally c-inconsistent. Then for some p it holds that $S \vdash p, -p$. By Lemma 39, $\{p, -p\} \subseteq Cn(S)$. By Definition 25(3b), $\{p, -p\}$ is AL-inconsistent, and so by monotonicity $Cn(S)$ is AL-inconsistent. But then by Definition 23-(2), S is AL-inconsistent, and so some finite $S' \subseteq S$ is AL-inconsistent. But since S is minimally inconsistent, it holds that $S' = S$. Consider any $s \in S$. Then $S \setminus \{s\} \cup \{s\}$ is inconsistent. By adjunction and finiteness of S there exists a formula $x \in \mathcal{L}$ that has exactly the same consequences as $S \setminus \{s\}$. Then by property (7) in Section 5.2 $\{x, s\}$ is AL-inconsistent, and so by Definition 25(3c) there exists a $y \in Cn(\{x\})$ such that $y = -s$. But then $S \setminus \{s\} \vdash -s$. □

Lemma 42. *For any $(AS, \mathcal{K})$ based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$, it holds that $\{p, q\}$ is AL-inconsistent iff $p \vdash -q$ for some $-q$.*

Proof. Suppose $\{p, q\}$ is AL-inconsistent. By Definition 25(3) $-q \in Cn\{p\}$ for some $-q$. By Lemma 39, $p \vdash -q$. Suppose $p \vdash -q$ for some $-q$. Then $-q \in Cn(\{p\})$ by Lemma 39. By monotonicity $-q \in Cn(\{p, q\})$. Furthermore, $q \in Cn(\{p, q\})$

by Definition 23-(1). Since q and $-q$ are contradictories, then by Definition 25(3)-a&b, $\{q, -q\}$ is AL-inconsistent. Hence $Cn(\{q, -q\}) = \mathcal{L}$. Then by Definition 23-(2) and monotonicity, $Cn(\{p, q\}) = \mathcal{L}$. □

Proposition 26. *Let $(AS, \mathcal{K})$ be based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Let Δ_1 be the c-SAF defined by $(AS, \mathcal{K})$ and $\leqslant'$, and Δ_2 the AL-c-SAF defined by $(AS, \mathcal{K})$ and $\leqslant'$. Then, for $T \in$ {complete, grounded, preferred, stable}, E is a T extension of Δ_1 iff E is a T extension of Δ_2.*

Proof. We first show that:

$$X \text{ is acceptable w.r.t. } E \text{ in } \Delta_1 \text{ iff } X \text{ is acceptable w.r.t. } E \text{ in } \Delta_2. \quad (\mathbf{1})$$

1.1) Firstly, if Y undermines X, then p $(= \texttt{Conc}(Y)) \in \bar{q}$, where $q \in \texttt{Prem}(X)$. By Definition 25(3)-b, $\{p, q\}$ is AL-inconsistent. By Lemma 42, $p \vdash -q$. Hence Y *ASPIC*$^+$-AL-undermines X.

1.2) Secondly, if Y *ASPIC*$^+$-AL-undermines X, then p $(= \texttt{Conc}(Y)) \vdash -q$, where $q \in \texttt{Prem}(X)$. Hence Y can be strictly continued into an argument Y' that concludes $-q$, and so Y' undermines X.

Suppose Y defeats X in Δ_1 ($Y \hookrightarrow_{\Delta_1} X$), and so $\exists Z \in E$, $Z \hookrightarrow_{\Delta_1} Y$. By 1.1), $Y \hookrightarrow_{\Delta_2} X$ and $Z \hookrightarrow_{\Delta_2} Y$. Hence the left to right half of 1) is shown.

Suppose $Y \hookrightarrow_{\Delta_2} X$, and so $\exists Z \in E$, $Z \hookrightarrow_{\Delta_2} Y$. By 1.2) there is a strict continuation Y' of Y s.t. Y' undermines X. By condition 1-iii) of $\preccurlyeq$ being reasonable (Definition 18), it remains the case that $Y' \not\prec X$, and so $Y' \hookrightarrow_{\Delta_1} X$. By the same reasoning, there is a strict continuation Z' of Z, s.t. $Z' \not\prec Y$ and so $Z' \hookrightarrow_{\Delta_1} Y$. Since Z' undermines Y then Z' undermines Y'. By condition 1-iii) of $\preccurlyeq$ being reasonable, $Z' \not\prec Y'$. Hence $Z' \hookrightarrow_{\Delta_1} Y'$. By Lemma 37, Z' is acceptable w.r.t. E, and by Proposition 10, $E \cup \{Z'\}$ is conflict free. Hence, since E is complete, $Z' \in E$. Hence the right to left half of 1) is shown.

Notice that (**1**) is shown in exactly the same way assuming the defeat definition of conflict free, except that in the final part of the above proof, we include the extra step of reasoning that since $E \cup \{Z'\}$ is conflict free under the attack definition, it trivially follows that $E \cup \{Z'\}$ is conflict free under the defeat definition (since the defeat relation is a subset of the attack relation).

Given (**1**), the main proposition now follows from the following:

$$E \text{ is conflict free in } \Delta_1 \text{ iff } E \text{ is conflict free in } \Delta_2. \quad (\mathbf{2})$$

Proof of (**2**) *under the attack definition of conflict free*: Suppose E is conflict free in Δ_1, E is not conflict free in Δ_2. Then $\exists Y, X \in E$ such that Y does not undermine X, Y *ASPIC*$^+$-AL-undermines X. By 1.2, there is a strict continuation Y' of Y s.t. Y' undermines X. Applying the same reasoning as for Z' above, $Y' \in E$, contradicting E is conflict free in Δ_1. Suppose E is conflict free in Δ_2, E is not conflict free in Δ_1. Then $\exists Y, X \in E$ such that Y does not *ASPIC*$^+$-AL-undermines X, Y undermines X, contradicting 1.1. Hence (**2**) is shown.

Proof of (**2**) *under the defeat definition of conflict free*: Suppose E is conflict free in Δ_1, E is not conflict free in Δ_2. Then $\exists Y, X \in E$ such that Y does not defeat X in Δ_1, Y defeats X in Δ_2. Given the latter, Y *ASPIC*$^+$-AL-undermines X on X', $Y \not\prec X'$. But then there is a strict continuation Y' of Y s.t. Y' undermines X on X', and by condition 1-iii) of $\preccurlyeq$ being reasonable, $Y' \not\prec X'$, and so Y' defeats X on X'. Applying the same reasoning as for Z' above, $Y' \in E$, contradicting E is conflict free in Δ_1.

Suppose E is conflict free in Δ_2, E is not conflict free in Δ_1. Then $\exists Y, X \in E$ such that Y does not defeat X in Δ_2, Y defeats X in Δ_1. Given the latter, Y undermines X on X', $Y \not\prec X'$. But then we immediately have that Y *ASPIC*$^+$-AL-undermines X on X' and so Y defeats X in Δ_2. Contradiction. □

Proposition 27. *Let $(AS, \mathcal{K})$ be based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Then A is a c-consistent premise minimal argument on the basis of $(AS, \mathcal{K})$ iff $(\texttt{Prem}(A), \texttt{Conc}(A))$ is an abstract logic argument on the basis of $(\Sigma, \leqslant')$.*

Proof. *Right to left*: let (X, p) be an *AL* argument. Then $X \rightarrow p \in \mathcal{R}_s$, and so A is a strict *ASPIC*$^+$ argument with $\texttt{Prem}(A) = X$, $\texttt{Conc}(A) = p$. A must be premise minimal since otherwise there is a strict *ASPIC*$^+$ argument A' for p with $X' = \texttt{Prem}(A')$, $X' \subset X$. But then by Lemma 39, $p \in Cn(X')$, contradicting the minimality of (X, p). Suppose for contradiction that A is not c-consistent. Then $X \vdash p, -p$, and so by Lemma 39, $\{p, -p\} \subseteq Cn(X)$. By Definition 25(3b), $\{p, -p\}$ is AL-inconsistent, and so $Cn(\{p, -p\}) = \mathcal{L}$. Then by Definition 23-(2) and monotonicity, $Cn(X) = \mathcal{L}$, so X is AL-inconsistent. Contradiction.

Left to right: let A be a c-consistent premise minimal *ASPIC*$^+$ argument with $\texttt{Prem}(A) = X$, $\texttt{Conc}(A) = p$ (i.e., $X \vdash p$). Then $X \subseteq \Sigma$ and $p \in Cn(X)$ (by Lemma 39), satisfying (1) and (3) of Definition 24. If $p \in Cn(X')$ for some $X' \subset X$, then $X' \rightarrow p \in \mathcal{R}_s$, and since $X' \subseteq \mathcal{K}_p$, there is an A' s.t. $\texttt{Prem}(A') = X'$, $\texttt{Conc}(A) = p$, contradicting A is premise minimal. Hence condition (4) of Definition 24 is satisfied. Suppose for contradiction that X is AL-inconsistent. By repeated application of adjunction to the formulae in $X \setminus \{\varphi\}$, for some $\varphi \in X$, we have that $Cn(X) = Cn(\{\phi, \varphi\}) = \mathcal{L}'$. By Definition 25(3c), $\exists \phi' \in Cn(\{\phi\})$, s.t. $\phi' \in \overline{\varphi}$, and so $\phi' = -\varphi$. By monotonicity, $-\varphi \in Cn(\{\phi, \varphi\})$. By Definition 23-(1), $\varphi \in Cn(\{\phi, \varphi\})$. Hence $\varphi, -\varphi \in Cn(X)$. By Lemma 39, $X \vdash \varphi$, $X \vdash -\varphi$; i.e., A is c-inconsistent. Contradiction. □

In the following lemma we introduce, for any argument $A \in \mathcal{A}$, the additional notation A_+ to denote any argument such that $\texttt{Prem}(A) \subseteq \texttt{Prem}(A_+)$ and $\texttt{Conc}(A) = \texttt{Conc}(A_+)$.

Lemma 43. *Consider any AT for which $\preccurlyeq$ is defined such that $\forall A, B, A_-$, if $A \not\prec B$ then $A_- \not\prec B$.*

1. *If A defeats B then A_- defeats B_+ for all A_- and B_+.*
2. *For all complete extensions E:*
   (a) *if $A \in E$ then $A_- \in E$ for all A_-;*
   (b) *if $B \notin E$ then $B_+ \notin E$ for all B_+.*

Proof. (1) If A defeats B based on a preference independent attack on a sub-argument B' of B, then A_- preference independent attacks and defeats B_+ on B'; else A preference dependent attacks and defeats B on B', $A \not\prec B'$. Since $A_- \not\prec B'$, A_- defeats B_+ (on B').

(2a) Let $A \in E$ and B defeat any A_-. Then B also defeats A by 1. Then there exists a C that defeats B, so A_- is acceptable with respect to A, so (since E is complete) $A_- \in E$.
(2b) Let $B \notin E$. Then there exists an A that defeats B, and no $C \in E$ s.t. $C \hookrightarrow A$. But then A defeats B_+ by 1, so B_+ is not acceptable with respect to E. □

Proposition 28. *Let Δ be the (c)-SAF $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ defined on the basis of an AT for which $\preccurlyeq$ is defined such that for any $A \in \mathcal{A}$, if $A \not\prec B$ then $A_- \not\prec B$.*

Let Δ^- be the premise minimal (c)-SAF $(\mathcal{A}^-, \mathcal{C}^-, \preccurlyeq^-)$ where:

- *$\mathcal{A}^-$ is the set of premise minimal arguments in $\mathcal{A}$.*
- *$\mathcal{C}^- = \{(X, Y) | (X, Y) \in \mathcal{C}, X, Y \in \mathcal{A}^-\}$.*
- *$\preccurlyeq^- = \{(X, Y) | (X, Y) \in \preccurlyeq, X, Y \in \mathcal{A}^-\}$.*

Then for $T \in$ {complete, grounded, preferred, stable}, E is a T extension of Δ iff E' is a T extension of Δ^-, where $E' \subseteq E$ and $\bigcup_{X \in E} \texttt{Conc}(X) = \bigcup_{Y \in E'} \texttt{Conc}(Y)$.

Proof.
$T =$ ***complete***:
1) Suppose E is a complete extension of Δ. We show that $E-$ is a complete extension of Δ^-, where $E- = E \cap \mathcal{A}^-$.
Note first that $E-$ is (attack) conflict-free by construction.

(i) Let $A \in E-$ and consider any $B \in \mathcal{A}^-$ that defeats A. Since $A \in E$, there exists a $C \in E$ that defeats B. But then (by Lemma 43(1)) for all C_- we also have that C_- defeats B. Since all such $C_- \in E$ (by Lemma 43(2)), all such C_- are in $E-$, and so A is acceptable with respect to $E-$.
(ii) Let $A \in \mathcal{A}^-$, $A \notin E-$. Then $A \notin E$, some B defeats A, $\neg\exists C \in E$, C defeats B. There exists a B_- that (by Lemma 43(1)) defeats A. Suppose for contradiction that A is acceptable w.r.t. $E-$. Then, $\exists C' \in E^-$, C' defeats B_-. By Lemma 43(1), C' defeats B. Since $E- \subseteq E$, $C' \in E$, contradicting $\neg\exists C \in E$, C defeats B. So A is not acceptable with respect to $E-$.

Given (i) and (ii), $E-$ is a complete extension of Δ^-.
2) Suppose E is a complete extension of Δ^-. We show that $E+$ is a T-extension of Δ, where $E+ = E \cup \{A_+ \in \mathcal{A} \mid A \in E$ and $\texttt{Prem}(A_+) \subseteq E\}$. Suppose E is a complete extension of Δ^-.
Note first that $E+$ is conflict-free by construction.

(i) For any $A \in E$ and any $B \in \mathcal{A}$ that defeats A, we have that some $B_- \in \mathcal{A}^-$ defeats A by Lemma 43(1), so some $C \in E$ defeats B_-. But then C also defeats B by Lemma 43(1). Since $C \in E+$, we have that A is acceptable with respect to $E+$.
(ii) Consider any $A_+ \in E+$, $A_+ \notin E$. Suppose $B \in \mathcal{A}$ defeats A_+. Then B defeats A_+ on some A' that is a premise in $\texttt{Prem}(A_+)$. By definition of $E+$ and sub-argument closure of E, $A' \in E$. By Lemma 43(1), $B_- \in \mathcal{A}^-$ defeats A', and B_- is defeated by some $C \in E$. Since C defeats B (by Lemma 43(1)) and $C \in E+$, A_+ is acceptable with respect to $E+$.
(iii) Consider finally any $A \in \mathcal{A}$, $A \notin E+$. Then no A_- is in E, so for all A_- there exists a B defeats A_-, $\neg\exists C \in E$, C defeats B. By Lemma 43(1) B also defeats A. Suppose for contradiction that A is acceptable w.r.t. $E+$, and so $\exists C' \in E+$, C' defeats B. Hence C' must be some C_+, where by Lemma 43-2 and construction of $E+$, $C \in E+$ and $C \in E$. By Lemma 43(1), C defeats B, contradicting $\neg\exists C \in E$, C defeats B.

By (i), (ii) and (iii), $E+$ is a complete extension of Δ.

$T =$ ***preferred***:

1) Suppose E is a preferred extension of Δ. Suppose for contradiction that $E-$ is not a preferred extension of $\Delta-$. We have shown that $E-$ is a complete extension of $\Delta-$. Hence there must be some $E' \supset E-$ that is a complete extension of $\Delta-$. We have shown that $E'+$ is a complete extension of Δ. It is easy to see by construction of $E-$ and $E'+$ that $E \subset E'+$, contradicting E is a preferred extension of Δ.

2) Suppose E is a preferred extension of $\Delta-$. Suppose for contradiction that $E+$ is not a preferred extension of Δ. We have shown that $E+$ is a complete extension of Δ. Hence there must be some $E' \supset E+$ that is a complete extension of Δ. We have shown that $E'-$ is a complete extension of $\Delta-$. It is easy to see by construction of $E'-$ and $E+$, that $E \subset E'-$, contradicting E is a preferred extension of $\Delta-$.

$T =$ ***grounded***:

1) Suppose E is the grounded extension of Δ. Suppose for contradiction that $E-$ is not the grounded extension of $\Delta-$. We have shown that $E-$ is a complete extension of $\Delta-$. Hence there must be some $E' \subset E-$ that is a complete extension of $\Delta-$. We have shown that $E'+$ is a complete extension of Δ. It is easy to see by construction of $E-$ and $E'+$ that $E'+ \subset E$, contradicting E is the grounded extension of Δ.

2) Suppose E is the grounded extension of $\Delta-$. Suppose for contradiction that $E+$ is not the grounded extension of Δ. We have shown that $E+$ is a complete extension of Δ. Hence there must be some $E' \subset E+$ that is a complete extension of Δ. We have shown that $E'-$ is a complete extension of $\Delta-$. It is easy to see by construction of $E'-$ and $E+$, that $E'- \subset E$, contradicting E is the grounded extension of $\Delta-$.

$T =$ ***stable***:

1) Let E be a stable (and so preferred) extension of Δ. Then $E-$ is a preferred extension of $\Delta-$. Suppose for contradiction that $E-$ is not a stable extension. Then $\exists B \in \mathcal{A}-$, $B \notin E-$, and B is not defeated by an argument in $E-$. Note that $B \in \mathcal{A}$. It cannot be that $B \in E$ since the fact that $B \in \mathcal{A}-$ would imply by construction of $E-$ that $B \in E-$. Since E is stable, some $C \in E$ defeats B. By Lemma 43(1) all C_- also defeat B, and by Lemma 43(2a) all such C_- are in E. By construction of $E-$ all such C_- are in $E-$, and so B is defeated by an argument in $E-$. Contradiction.

2) Let E be a stable (and so preferred) extension of $\Delta-$. Then $E+$ is a preferred extension of Δ. Suppose for contradiction that $E+$ is not a stable extension. Then $\exists B \in \mathcal{A}$, $B \notin E+$, and B is not defeated by an argument in $E+$. Since $E+$ is preferred, B is defeated by a $C \in \mathcal{A} \setminus E+$, where C is not defeated by an argument in $E+$.

C defeats B on some $\varphi \in \mathtt{Prem}(B)$. Note that $\varphi \in \mathcal{A}-$. If $\varphi \in E$ then, since φ acceptable w.r.t. E, C is defeated by some argument in E. But since $E \subseteq E+$ this contradicts that C is not defeated by an argument in $E+$. If $\varphi \notin E$ then, since E is a stable extension of $\Delta-$, we have that φ is defeated by a $D \in E$, but then D also defeats B. Since $D \in E+$ this contradicts that B is not defeated by an argument in $E+$.

Finally, we clearly have for any E and $E-$ that $\mathtt{Conc}(E) = \mathtt{Conc}(E-)$, and likewise for any E and $E+$. Then the proposition follows from (1) and (2) for each of the above semantics. □

Corollary 29. *Given Δ and Δ^- as defined in Proposition 28*:

1. *φ is a T credulously (sceptically) justified conclusion of Δ iff φ is a T credulously (sceptically) justified conclusion of Δ^-.*
2. *Δ^- satisfies the postulates* closure under strict rules, direct consistency, indirect consistency *and* sub-argument closure.

Proof. 1) and *closure under strict rules*, *direct consistency*, and *indirect consistency* immediately follow from Proposition 28. For *sub-argument closure* expressed in Theorem 12, note that the proof of this theorem appeals to Lemma 35 which can straightforwardly be seen to apply to Δ^-. The proof also depends on any sub-argument of $A \in E$ not being in conflict with any argument in E. This immediately follows for $E-$ in the proof of Proposition 28, given that $E- \subseteq E$. □

Proposition 30. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be defined by an AL argumentation theory, where $\preccurlyeq$ is defined under the weakest or last link principles, based on the set comparison $\trianglelefteq_{\mathtt{Eli}}$. Then $\forall A, B \in \mathcal{A}$, $\forall A_- \in \mathcal{A}$, if $A \not\prec B$ then $A_- \not\prec B$.*

Proof. Since all arguments are strict continuations of ordinary premises, the last and weakest link principles are evaluated in the same way. Suppose $A \not\prec B$. Then $\mathtt{Prem}(A) \not\trianglelefteq_{\mathtt{Eli}} \mathtt{Prem}(B)$. That is to say, it is not the case that $\exists X \in \mathtt{Prem}(A)$ s.t. $\forall Y \in \mathtt{Prem}(B)$, $X \leqslant Y$, i.e., $\forall X \in \mathtt{Prem}(A)$, $\exists Y \in \mathtt{Prem}(B)$ s.t. $X \not\leqslant Y$. Since $\mathtt{Prem}(A_-) \subseteq \mathtt{Prem}(A)$, it trivially follows that $\forall X \in \mathtt{Prem}(A_-)$, $\exists Y \in \mathtt{Prem}(B)$ s.t. $X \not\leqslant Y$, i.e., $A_- \not\prec B$. □

Proposition 32. *Let Δ be the c-SAF based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leqslant')$. Then for any complete extension E of Δ: $S = \{\phi | \phi \in \mathtt{Prem}(A)$, $A \in E\}$ is AL-inconsistent iff $S' = Cl_{\mathcal{R}_s}(\{\mathtt{Conc}(A) | A \in E\})$ is inconsistent.*

Proof. *Left to right*: if S is AL inconsistent then $\varphi, -\varphi \in Cn(S)$ for any φ. By definition of $\mathcal{R}_s$, for some $T, T' \subseteq S$ there exist rules $T \rightarrow \varphi$ and $T' \rightarrow -\varphi$ in $\mathcal{R}_s$. Since E is closed under sub-arguments and premises are sub-arguments, $\{\mathtt{Conc}(A) | A \in E\}$ includes T and T'. Hence $\varphi, -\varphi \in S'$. That is, S' is inconsistent.

Right to left: If S' is inconsistent then $\varphi, -\varphi \in S'$ for some φ. Since E is closed under sub-arguments, $S \subseteq S'$, and so $S \vdash \varphi$ and $S \vdash -\varphi$. By Definition 25(3b), $\{\varphi, -\varphi\}$ is AL-inconsistent, so $Cn(\{\varphi, -\varphi\}) = \mathcal{L}$. But since $\{\varphi, -\varphi\} \subseteq Cn(S)$ and $Cn(Cn(S)) = Cn(S)$, we have by monotonicity of Cn that $Cn(S) = \mathcal{L}$ so S is AL-inconsistent. □

A.6. Proofs for Section 5.3

Theorem 34. *Let $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$ be a c-SAF corresponding to a default theory Γ, and for any $\Sigma \subseteq \Gamma$, let $\mathtt{Args}(\Sigma) \subseteq \mathcal{A}$ be the set of all arguments with premises taken from Σ. Then*:

1) *If Σ is a preferred subtheory of Γ, then $\mathtt{Args}(\Sigma)$ is a stable extension of $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$.*
2) *If E is a stable extension of $(\mathcal{A}, \mathcal{C}, \preccurlyeq)$, then $\bigcup_{A \in E} \mathtt{Prem}(A)$ is a preferred subtheory of Γ.*

Proof of 1). Firstly, we show that $\mathtt{Args}(\Sigma)$ is conflict free. Since Σ is consistent, $\Sigma \nvdash_c \alpha, \neg\alpha$ for any α. Suppose for contradiction that $\mathtt{Args}(\Sigma)$ is not conflict free, in which case $\exists X, Y \in \mathtt{Args}(\Sigma)$ s.t. $\mathtt{Conc}(X) = \alpha$, $\neg\alpha \in \mathtt{Prem}(Y)$. But then since every such argument is obtained by applying the strict rules encoding all classical inferences to Σ, this implies $\Sigma \vdash_c \alpha, \neg\alpha$. Contradiction.

We now show that for any $Y \in \mathcal{A} \setminus \mathtt{Args}(\Sigma)$, $\exists X \in \mathtt{Args}(\Sigma)$ s.t. X defeats Y. Consider any such Y. Then $\exists \gamma \in \mathtt{Prem}(Y)$, $\gamma \notin \Sigma$. By construction, $\Sigma = \Sigma_1 \cup \cdots \cup \Sigma_n$ such that for $i = 1 \ldots n$, $\Sigma_1 \cup \cdots \cup \Sigma_i$ is a maximal consistent subset of $\Gamma_1, \ldots, \Gamma_i$. Hence, suppose $\gamma \in \Gamma_j$ for some $j = 1 \ldots n$. Then $\Sigma_1 \cup \cdots \cup \Sigma_j \cup \{\gamma\} \vdash_c \bot$. Hence $\Sigma_1 \cup \cdots \cup \Sigma_j \vdash_c \neg\gamma$. Hence, $\exists X \in \mathtt{Args}(\Sigma_1 \cup \cdots \cup \Sigma_j)$ s.t. $\mathtt{Conc}(X) = \neg\gamma$, and so $X \rightharpoonup Y$. Since $\gamma \in \Gamma_j$, and all premises in X are in Γ_i, $i \leqslant j$ (i.e., every premise in X is greater or equal to γ) then $\mathtt{Prem}(Y) \trianglelefteq_{\mathtt{Eli}} \mathtt{Prem}(X)$, and so by the weakest or last link principle, $X \not\prec Y$. Hence $X \hookrightarrow Y$.

Proof of 2). Firstly, we show that $\bigcup_{A \in E} \mathtt{Prem}(A)$ must be consistent. Suppose for contradiction that $\exists X, Y \in E$ s.t. $\mathtt{Prem}(X) \cup \mathtt{Prem}(Y) \vdash_c \bot$. Let $\{\alpha_1, \ldots, \alpha_m\}$ be a minimal (under set inclusion) subset of $\mathtt{Prem}(X) \cup \mathtt{Prem}(Y)$ s.t. $\alpha_1, \ldots, \alpha_m \vdash \bot$. Hence, $\alpha_1, \ldots, \alpha_{m-1} \vdash \neg\alpha_m$. Since E is stable and so complete, then by sub-argument closure (Theorem 12), $\{A_1, \ldots, A_m\} \subseteq E$, where for $i = 1 \ldots m$, $\mathtt{Prem}(A_i) = \{\alpha_i\}$. By Lemma 37, if $\{A_1, \ldots, A_m\} \subseteq E$, where E is a complete extension, then any strict continuation of $\{A_1, \ldots, A_m\}$ is acceptable w.r.t. E, and so in E. Hence $A \in E$ where A concludes $\neg\alpha_m$. Hence, $A \rightharpoonup A_m$, contradicting E is conflict free.

Next, let $E_1, \ldots, E_n$ be the partition of $\mathtt{Form}(E)$ s.t. for $i = 1 \ldots n$, E_i is a (possibly) empty subset of Γ_i in the stratification $\Gamma_1, \ldots, \Gamma_n$ of Γ. Suppose for contradiction that $\mathtt{Form}(E)$ is not a preferred subtheory. Then, for some i, for $k = 1 \ldots i-1$, $E_1, \ldots, E_k$ is a maximal consistent subset of $\Gamma_1, \ldots, \Gamma_{i-1}$, and $\exists \alpha \in \Gamma_i$ s.t. $\alpha \notin E_i$, and $E_1 \cup \cdots \cup E_{i-1} \cup E_i \cup \{\alpha\} \nvdash \bot$. Hence, $\exists Y \in \mathcal{A}$, $\mathtt{Prem}(Y) = \{\alpha\}$, $Y \notin E$. By assumption of E being a stable extension, $\exists X \in E$, $X \hookrightarrow Y$. Since $E_1 \cup \cdots \cup E_{i-1} \cup E_i \cup \{\alpha\} \nvdash \bot$, then $E_1 \cup \cdots \cup E_{i-1} \cup E_i \nvdash \neg\alpha$, and so it must be that some $\beta \in \mathtt{Prem}(X)$ is in E_j, $j > i$; i.e., $\beta \in \mathtt{Prem}(X)$, $\mathtt{Prem}(Y) = \{\alpha\}$, and $\beta < \alpha$. Hence $\mathtt{Prem}(X) \trianglelefteq_{\mathtt{Eli1}} \mathtt{Prem}(Y)$, $\mathtt{Prem}(Y) \not\trianglelefteq_{\mathtt{Eli1}} \mathtt{Prem}(X)$, and so $X \prec Y$ under the weakest or last link principle, contradicting $X \hookrightarrow Y$. □

References

[1] L. Amgoud, Five weaknesses of ASPIC+, in: Proc. 14th International Conference on Information Processing and Management of Uncertainty in Knowledge Based-Systems (IPMU'12), 2012, pp. 122–131.
[2] L. Amgoud, P. Besnard, Bridging the gap between abstract argumentation systems and logic, in: Proc. 3rd International Conference on Scalable Uncertainty (SUM'09), 2009, pp. 12–27.
[3] L. Amgoud, P. Besnard, A formal analysis of logic-based argumentation systems, in: Proc. 4th International Conference on Scalable Uncertainty (SUM'10), 2010, pp. 42–55.
[4] L. Amgoud, C. Cayrol, Inferring from inconsistency in preference-based argumentation frameworks, International Journal of Automated Reasoning 29 (2) (2002) 125–169.
[5] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, Annals of Mathematics and Artificial Intelligence 34 (1–3) (2002) 197–215.
[6] L. Amgoud, G. Vesic, Generalizing stable semantics by preferences, in: P. Baroni, F. Cerutti, M. Giacomin, G.R. Simari (Eds.), Computational Models of Argument. Proceedings of COMMA 2010, pp. 39–50.
[7] L. Amgoud, S. Vesic, Repairing preference-based argumentation frameworks, in: Proc. 21st International Joint Conferences on Artificial Intelligence, 2009, pp. 665–670.
[8] L. Amgoud, S. Vesic, Handling inconsistency with preference-based argumentation, in: Proc. 4th International Conference on Scalable Uncertainty Management (SUM'10), 2010, pp. 56–69.
[9] L. Amgoud, S. Vesic, A new approach for preference-based argumentation frameworks, Annals of Mathematics and Artificial Intelligence 63 (2) (2011) 149–183.
[10] T.J.M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, Journal of Logic and Computation 13 (3) (2003) 429–448.
[11] T.J.M. Bench-Capon, P.E. Dunne, Argumentation in artificial intelligence, Artificial Intelligence 171 (2007) 10–15.
[12] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artificial Intelligence 128 (2001) 203–235.
[13] P. Besnard, A. Hunter, Elements of Argumentation, MIT Press, 2008.
[14] F.J. Bex, S. Modgil, H. Prakken, C. Reed, On logical specifications of the argument interchange format, Journal of Logic and Computation, http://dx.doi.org/10.1093/logcom/exs033, 2012.
[15] A. Bondarenko, P.M. Dung, R.A. Kowalski, F. Toni, An abstract, argumentation-theoretic approach to default reasoning, Artificial Intelligence 93 (1997) 63–101.

[16] G. Brewka, Preferred subtheories: An extended logical framework for default reasoning, in: Proc. 11th International Joint Conference on Artificial intelligence, 1989, pp. 1043–1048.
[17] G. Brewka, Nonmonotonic Reasoning: Logical Foundations of Commonsense, Cambridge University Press, 1991.
[18] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, Artificial Intelligence 171 (5–6) (2007) 286–310.
[19] M. Caminada, W. Carnielli, P.E. Dunne, Semi-stable semantics, Journal of Logic and Computation 22 (5) (2012) 1207–1254.
[20] C. Cayrol, On the relation between argumentation and non-monotonic coherence-based entailment, in: Proc. 14th International Joint Conference on Artificial intelligence, 1995, pp. 1443–1448.
[21] C. Cayrol, V. Royer, C. Saurel, Management of preferences in assumption-based reasoning, in: Proc. 4th International Conference on Information Processing and Management of Uncertainty in Knowledge Based-Systems (IPMU'92), 1992, pp. 13–22.
[22] C. Chesnevar, J. McGinnis, S. Modgil, Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, S. Willmott, Towards an argument interchange format, The Knowledge Engineering Review 21 (4) (2006) 293–316.
[23] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (2) (1995) 321–358.
[24] M.L. Ginsberg, AI and nonmonotonic reasoning, in: D. Gabbay, C.J. Hogger, J.A. Robinson (Eds.), Handbook of Logic in Artificial Intelligence and Logic Programming, Clarendon Press, Oxford, 1994, pp. 1–33.
[25] T.F. Gordon, H. Prakken, D.N. Walton, The Carneades model of argument and burden of proof, Artificial Intelligence 171 (2007) 875–896.
[26] N. Gorogiannis, A. Hunter, Instantiating abstract argumentation with classical logic arguments: Postulates and properties, Artificial Intelligence 175 (2011) 1479–1497.
[27] G. Governatori, M.J. Maher, An argumentation-theoretic characterization of defeasible logic, in: Proc. 14th European Conference on Artificial Intelligence, 2000, pp. 469–473.
[28] S. Kaci, Refined preference-based argumentation frameworks, in: P. Baroni, F. Cerutti, M. Giacomin, G.R. Simari (Eds.), Computational Models of Argument. Proceedings of COMMA 2010, pp. 299–310.
[29] R.A. Kowalski, F. Toni, Abstract argumentation, Artificial Intelligence and Law 4 (3) (1996) 275–296.
[30] F. Lin, Y. Shoham, Argument systems: A uniform basis for nonmonotonic reasoning, in: Proc. 1st International Conference on Principles of Knowledge Representation and Reasoning (KR-89), 1989, pp. 245–255.
[31] R.P. Loui, Defeat among arguments: A system of defeasible inference, Computational Intelligence 2 (1987) 100–106.
[32] H. Mercier, D. Sperber, Why do humans reason? Arguments for an argumentative theory, Behavioral and Brain Sciences 34 (2) (2011) 57–747.
[33] S. Modgil, Reasoning about preferences in argumentation frameworks, Artificial Intelligence 173 (9–10) (2009) 901–934.
[34] S. Modgil, M. Caminada, Proof theories and algorithms for abstract argumentation frameworks, in: I. Rahwan, G. Simari (Eds.), Argumentation in AI, Springer-Verlag, 2009, pp. 105–129.
[35] S. Modgil, H. Prakken, Reasoning about preferences in structured extended argumentation frameworks, in: P. Baroni, F. Cerutti, M. Giacomin, G.R. Simari (Eds.), Computational Models of Argument. Proceedings of COMMA 2010, pp. 347–358.
[36] S. Modgil, H. Prakken, Revisiting preferences and argumentation, in: International Joint Conference on Artificial Intelligence (IJCAI 2011), 2011, pp. 1021–1026.
[37] J.L. Pollock, Defeasible reasoning, Cognitive Science 11 (1987) 481–518.
[38] J.L. Pollock, Justification and defeat, Artificial Intelligence 67 (1994) 377–408.
[39] J.L. Pollock, Cognitive Carpentry. A Blueprint for How to Build a Person, MIT Press, Cambridge, MA, 1995.
[40] H. Prakken, An abstract framework for argumentation with structured arguments, Argument and Computation 1 (2) (2010) 93–124.
[41] H. Prakken, Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and dungean semantics, Artificial Intelligence and Law 20 (1) (2012) 57–82.
[42] H. Prakken, Some reflections on two current trends in formal argumentation, in: Logic Programs, Norms and Action. Essays in Honour of Marek J. Sergot on the Occasion of his 60th Birthday, Springer, Berlin/Heidelberg, 2012, pp. 249–272.
[43] H. Prakken, S. Modgil, Clarifying some misconceptions on the ASPIC+ framework, in: B. Verheij, S. Woltran, S. Szeider (Eds.), Computational Models of Argument. Proceedings of COMMA 2012, pp. 442–453.
[44] H. Prakken, G. Sartor, Argument-based extended logic programming with defeasible priorities, Journal of Applied Non-Classical Logics 7 (1997) 25–75.
[45] I. Rahwan, G. Simari (Eds.), Argumentation in AI, Springer-Verlag, 2009.
[46] G.R. Simari, R.P. Loui, A mathematical treatment of defeasible argumentation and its implementation, Artificial Intelligence 53 (1992) 125–157.
[47] B. van Gijzel, H. Prakken, Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation, Argument and Computation 1 (2012) 21–47.
[48] G.A.W. Vreeswijk, Abstract argumentation systems, Artificial Intelligence 90 (1997) 225–279.
[49] D.N. Walton, Argument Schemes for Presumptive Reasoning, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.
[50] Y. Wu, Between argument and conclusion. Argument-based approaches to discussion, inference and uncertainty, Chapter 6 in Doctoral Dissertation, University of Luxemburg, 2012.